ISOLATION,CHARACTERIZATION AND CHROMOSOMAL MAPPING OF HUMAN 56 kDa SELENIUM BINDING PROTEIN

BY

PETER,WEI GONG CHANG

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF PHILOSOPHY IN BIOCHEMISTRY

DECEMBER 1996

DIVISION OF BIOCHEMISTRY

THE CHINESE UNIVERSITY OF HONG KONG

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my supervisors, Prof. K. P. Fung and Prof. Mary M. W. Waye for their patient teaching, professional guidance and encouragement.

I wish to express my gratitude to Prof. C. C. Liew and Prof. C. Y. Lee for their insight and comments.

I am truly grateful to Dr. Stephen Tsui for his valuable advice and helpful suggestions.

I express my thanks to every member in the Biotechnology Lab, Rm 414 Basic Medical Sciences Building, who have rendered help in my study.

This work was supported by UGC Grants: CUHK 16/93 M and CUHK 418 /95 M.

ABSTRACT

The human heart cDNA sequencing project is a collaborative project between The Chinese University of Hong Kong and University of Toronto. During the course of my research studies, a novel cDNA clone coding for a putative human selenium binding protein was obtained from a human heart cDNA library. This full-length human 56 kDa selenium binding protein (hSP56) cDNA clone, which is the human homolog of mouse 56 kDa selenium binding protein is 1440bp long and has an open reading frame encoding 472 amino acids. The calculated molecular weight is 52.25 kDa and the estimated isoelectric point is 6.13. Using Northern blot hybridization, this 56 kDa selenium binding protein has been found to be expressed in mouse heart with an intermediate level between those found in liver / lung / kidney and intestine. hSP56 has also successfully been expressed in *E. coli* using the expression vector-pAED4. Under the conditions used, the hybridization efficiency was approximately 80% for the probe we used. The hSP56 gene

has been mapped to human chromosome by fluorescent *in situ* hybridization (FISH). The result shows that hSP56 gene is located at chromosome 1q21-22. As part of a collaborative study, I have sequenced 553 human heart cDNA clones as part of my project. I assigned putative identities of these clones, set up the database and catalogued the results.

Notes:

(1) Accession No. of hSP56 in GenBank: U29091

(2) Part of the results in this thesis were published as below:

Peter W. G. Chang, Stephen K. W. Tsui, Choong-chin Liew, Cheuk-yu Lee, Mary M. Y. Waye, and Kwok-pui Fung. (1997): Isolation, Characterization, and Chromosomal Mapping of a Novel cDNA Clone Encoding Human Selenium Binding Protein. *Journal of Cellular Biochemistry* 64: 217-224.

CONTENTS

CHAPTER 2 MATERIALS AND METHODS

ABBREVIATIONS

A.L.F	**Automatic Laser Fluorescent**
AP56	**Acetaminophen binding protein**
bp	**Base pair (s)**
BLAST	**Basic Local Alignment Search Tool**
Da	**Dalton**
dCTP	**Deoxycytosine 5'-triphosphate**
ddNTP	**2'3'-dideoxynucleotide 5'-triphosphate**
DMSO	**Dimethylsulfoxide**
dNTP	**Deoxynucleosine 5'-phosphate**
E.coli	***Escherichia coli***
EDTA	**Ethylenediaminetetracetic acid**
E. mail	**Electronic mail**
EMBL	**European Molecular Biology Laboratory**
EST	**Expressed sequence tag**
EtBr	**Ethidium bromide**
hSP56	**Human 56 kDa selenium binding protein**
IPTG	**Isopropyl-1-thio-β-D-galactopyranoside**
kb	**Kilobase (s)**
LB	**Luria-Bertani**

LBA	**Luria-Bertani-ampicillin**
LBAC	**Luria-Bertani-ampicillin-chloramphenicol**
LBC	**Luria-Bertani-chloramphenicol**
MOPS	**Morpholinopropanesulfonic acid**
mRNA	**Messenger RNA**
mSP56	**Mouse 56 kDa selenium binding protein**
MW	**Molecular weight**
NCBI	**National Center of Biotechnology Information**
ORF	**Open reading frame**
PAGE	**Polyacrylamide gel electrophoresis**
PCR	**Polymerase chain reaction**
pfu	**Plaque-forming unit**
SDS	**Sodium dodecyl sulfate**
SM	**Suspension medium**
SP	**Selenium binding protein**
SSC	**Standard saline citrate**
TEMED	**N, N, N', N'-tetramethylethylenediamine**
TM	**Annealing temperature**
UTR	**Untranslated region**
Xgal	**5-bromo-4-chloro-3-indolyl-β-D-galactoside**

Chapter 1

Introduction

1.1 General introduction

In this thesis, I report the partial sequencing of 553 cDNA clones which I randomly selected directly from a cDNA library. I will also describe the cloning, sequencing, tissue distribution, gene expression and chromosomal mapping of a novel human gene which we named as the human selenium binding protein 56 (hSP56).

The partial sequencing of randomly selected cDNA clones directly from a cDNA library has been explored by Putney and co-workers in 1983 (Putney *et al.,* 1983). Using this approach to generate expressed sequence tags (ESTs) has been shown to be a rapid and efficient way of establishing a detailed profile of genes expressed in various tissues (Adams *et al.,* 1991; Adams *et al.,* 1992; Adams *et al.,* 1993; Adams *et al.,* 1994; Collet *et al.,* 1994; Hwang *et al.,* 1994; Khan *et*

al., 1992; Liew *et al.*, 1994; Matoba *et al.*, 1994; Matsubara *et al.*, 1993; Okubo *et al.*, 1992; Park *et al.*, 1993; Sikela *et al.*, 1993; Tsui *et al.*, 1994; Waye *et al.*, 1995). ESTs are sequences from single sequencing runs of at least 150 base pairs with an accuracy of approximately 97% (Adams *et al.*, 1991). This approach can also be used to discover new genes on a large scale. The power and utility of EST data in the study of human diseases have also been recently manifested in the discovery of a human *mutL* homologue, which is the "Molecule of the Year" in 1994 (Koshland, 1994), putatively involved in hereditary colon cancer (Papadopoulos *et al.*, 1994). The isolation and sequencing of this homologue were tremendously facilitated by the prior availability of EST-tagged clones corresponding to the gene of interest. Information from single-pass sequencing of cDNA clones has also been used in many other applications, including the generation of physical maps of chromosomes (Chae *et al.*, 1994; Khan *et al.*, Wilcox *et al* 1991) and chromosomal assignment (Durkin *et al.*, 1992; Polymeropoulos *et al.*, 1992).

This thesis will be divided into 4 chapters. In chapter 1, the background information relevant to the thesis will be provided. Firstly, the genome project, the human adult and fetal heart cDNA library will be introduced. Then the background of the project will be presented. In addition, the background knowledge on the role of selenium will be described. The various types of selenium binding proteins will be reviewed. Finally, the introduction will be focused on an important protein: mouse 56kDa selenium-binding protein. The following three chapters will describe the study of hSP56. Firstly, in chapter 2, the materials and methods of study of hSP56 gene will be reported in detail. This chapter will present the method by computer analysis of hSP56 gene and the use of the computer for comparing the hSP56 with other related proteins such as the mSP56 gene. Then the study will show the analysis of hSP56 gene at RNA level and the cloning of hSP56 into an expression vector. The results may provide the information of the expression characterization of the hSP56 gene. Finally, I will describe the chromosomal mapping of hSP56. In chapter 3, I will report the results of sequence analysis, Northern blot

hybridization, cloning and expression of hSP56 gene and the chromosomal mapping result. Also I will show the sequencing result of 553 clones from the human adult heart cDNA library and the classification of these clones. Lastly, in chapter 4, I will discuss the results of those 553 clones and the possible roles of hSP56 and the future prospects for the study of hSP56 gene.

1.2 Human genome project

The Human Genome Project (HGP) was initiated within the United States Department of Energy (DOE) by Charles De Lisi from the Los Alamos Laboratory, New Mexico. A bill to fund the HGP was subsequently introduced by Mexico Senator Domenici (Rechsteiner *et al.,* 1991). The HGP first took clear form in February 1988, with the release of the National Research Council (NRC) report *Mapping and Sequencing the Human Genome* (Olson *et al.,* 1993). In 1990, the Human Genome programs of the National Institutes of Health (NIH) and the DOE developed a joint research plan with specific goals for the first 5 years (1991-95) of the U.S. HGP (Collins *et al.,* 1993). The initial set of goals are constructing detailed human genetic maps, improving physical maps of the human genome and the genomes of certain model organisms, developing improved technology for DNA sequencing and information handling and defining the most urgent set of ethical, legal and social issues associated with the acquisition and use of large amount of genetic information (Collins *et al.,* 1993).

In order to construct a high-resolution map of the human genome, a new choice of landmarks called sequence-tagged sites (STS) is required (Olson *et al.,* 1989; Olson *et al.,* 1993). An STS is simply a short, unique sequence of DNA that can be amplified via PCR. In addition, STSs can be described in an electronic database in a form that makes them experimentally accessible in any laboratory. The most critical aspect of an STS description is the DNA sequence of the two primers. Such a high-resolution map will require approximately 30,000 equally spaced STSs to achieve a 100 kb resolution (Olson *et al.,* 1989; Polymeropoulos *et al.,* 1992). Expressed sequence tags that are obtained from random sequencing of human cDNA libraries can be converted into STSs (Durkin *et al.,* 1992; Polymeropoulos *et al.,* 1992) and become useful resource for human genome mapping (Adams *et al.,* 1991; Wilcox *et al.,* 1991).

1.3 Human adult heart cDNA library

The human adult heart cDNA library used in this project was purchased from Clontech (Liew *et al.,* 1994). The mRNA was prepared from a human adult male heart, including some aorta region. The library was constructed with lambda gt11 as the cloning vector and EcoRI as the cloning site. It contains 1.27 X 10^6 independent clones. The range of insert sizes is 0.4 to 3.4 kb and the average insert size is 0.9 kb. This random-primed cDNA library was chosen because it is more informative in identifying genes and constructing a useful EST database than sequencing from the ends of full-length cDNAs (which contain non-coding 5' and 3' untranslated sequences) would be (Adams *et al.,* 1991).

1.4 Human fetal heart cDNA library

The human fetal heart cDNA library used in this project was obtained from Dr. C. C. Liew's Lab (Ontario, Canada). 6-12 week-old human fetal hearts were used to prepare the mRNA for the cDNA library. The library was constructed in the expression vector lambda gt22 as the cloning vector and *NotI- SalI* as the cloning site. The library was found to be of adequate complexity, containing greater than 1X 10^6 independent clones. Of 60 clones selected at random, average insert size was approximately 1.5 kb, with a range of 400 bp to greater than 30000 bp [Hwang *et al.,* 1994]. The information generated by EST sequencing of the human fetal heart library has at least twofold significance. First, the cataloguing of genes expressed in the developing heart will significantly advance the field of molecular cardiology [Liew *et al.,* 1994]. The second major impact of such sequencing will lie in its contribution to the human genome project. It is estimated that by 1998, at a time when human genomic sequencing will still be in its infancy, EST sequencing will already have identified and completely sequenced with an average three-fold redundancy the cDNAs of most genes expressed in the human body.

1.5 Sequencing of a human heart cDNA clone

Single-pass sequencing of randomly selected cDNA clones to generate sequence tags (ESTs) has been widely used to identify novel genes and to study gene expression in a variety of tissues (Adams *et al.,* 1991). Our human heart cDNA sequencing project is a collaborative project between The Chinese University of Hong Kong and University of Toronto (Liew *et al.*, 1993). It is a large scale project that aims at sequencing thousands of human heart cDNAs using the automatic sequencer.

I have used a commercially available human heart cDNA library and have randomly picked cDNA clones for this sequence analysis. The ESTs from the human heart cDNA library is a short DNA sequence which represents an expressed gene (Adams *et al.*, 1991). The study of this project will help to understand the gene expression profiles of human heart at different developmental stages and to identify novel genes which are specific to human heart (Liew

et al., 1994). It might also help to understand the mechanisms that govern the cells, tissues, and organs of the human body and to discern the basis of genetic diseases. The research in these novel genes may have significance in clinical studies. For example, a cDNA clone that is similar to some non-human cDNAs will be placed with special attention. This gene may be the human homology of , or related to, the non-human genes that were matched.

I have chosen the human heart tissue because cardiovascular disease is the most frequent cause of death in adult life in industrialised societies, and is increasingly important in developing countries. In early life minor congenital abnormalities affect 1 in 100 live births and more serious abnormalities approximately 1 in 500 (Edwards *et al.*, 1996). Our genetic approach of cardiology will complement those mainly focused on the consequences of malformation, metabolic disorder, auto-immune damage and degenerative disease. The heart is not only composed of cardiocytes but also other cell types, such as fibroblasts and endothelial

cells. It also contains neurovascular components such as nerve cells, smooth muscles cells, etc. As a result, certain abnormal genetic signals leading to the development of heart diseases may be revealed by the sequencing of a cardiovascular cDNA library (Michel and Driscoll, White and Lalouel, 1988).

Both human fetal heart and human adult heart cDNA libraries were sequenced in our research team study. I have chosen both human adult and fetal heart cDNA library in my study project, because of the fact that screening human fetal heart cDNA library would help to characterize cardiac development at a transcriptional level and that it would increase the probability of detecting key genes involved in developmental and disease processes. The composition of ESTs obtained from the fetal heart library differed somewhat from that of the adult heart library previously reported (Liew *et al.*, 1994), as evidenced by differences in the proportions of known ESTs in several functional categories. Most striking was the relative abundance of transcripts representing basic

transcriptional and translational (TC/TL) apparatus, as well as transcripts representing signal transduction or cell regulatory proteins in the fetal heart, compared with the relative lack of such proteins in the adult heart. In contrast, the fetal heart appears to express lower levels of transcripts representing contractile elements than the adult heart (Hwang *et al*,1995).

One of the disadvantages of random cDNA sequencing is that highly expressed genes will be sequenced repeatedly while the rarely expressed genes may not be included in the data base unless a large number of cDNA clones are sequenced (Okubo *et al.*,1992).

1.6 Background knowledge on the role of selenium

Basic knowledge on the role of selenium in biology and human physiology is accumulating at a rapid rate (Robberecht and Deelstra, 1994). The first description of selenium was published by Berzelius, who obtained a reddish precipitate in 1817 while burning pyrite in a sulfuric acid plant. Upon analysis of this precipitate, it turned out to be a new element. He called the element after Selene, the Greek goddess of the moon, as tellurium, which was discovered earlier was named after the goddess of earth, Tellus (Nebergall *et al.*, 1968).

Selenium is an element of subgroup VIA of the Periodic Table. This subgroup comprises the non-metals oxygen (O) and sulfur (S), the metalloid selenium (Se) and the metals tellurium (Te) and polonium (Po). Among these elements the close relationship in chemical properties of selenium and sulfur is most striking. Due to this chemical similarity, no discrimination between sulfur and selenium is made in

certain biological processes and incorporation of selenium and sulfur into proteins occurs in parallel (Hurd *et al.,* 1964). There is a wide distribution of selenium throughout the earth's crust, similar to that of sulfur. The concentration of selenium on earth averages 0.09 ppm but shows a large range. Concentration of selenium in soil range from less than 0.1 ppm in the selenium deficient areas of New Zealand to as high as 1200 ppm in the organic soil in Ireland (Frankenberger *et al.,* 1994).

During half a century, the public image of selenium gradually evolved from that of a highly toxic and carcinogenic element to an essential trace element with antioxidant and anticarcinogenic properties. The role of selenium as a micronutrient has been well established. The average daily selenium intake in the 1980s was 40-50 μg/day. However, most of the time the selenium dietary intake was below the lower limit of the safe and adequate intake for selenium as defined by the U.S. National Academy of Sciences (recommended daily allowance (RDA) of 50

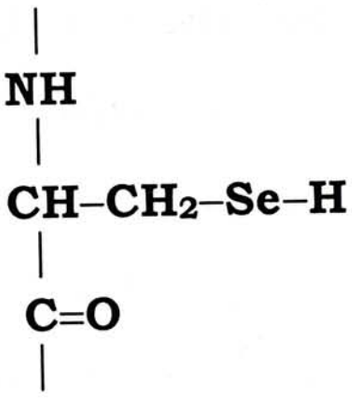
μg/day) (Burk. 1995). Deficiency of selenium has been associated with reduced life span, anemia, liver necrosis and immunological disorders. Selenium deficiency also causes structural deformation of sperm tail in rats and muscular dystrophy in lambs. Dietary supplementation of selenium have been shown to reduce the incidence of cancer of colon, liver, intestine, skin, and of mammary glands in animal studies (Bansal *et al.,* 1990).

Selenium deficiency has also been observed in human. This selenium-responsive disease, known as Keshan disease, is a juvenile cardiomyopathy occurring in the Heilongjiang province of the People's Republic of China. It is the major cause of the death for young children. Since the administration of 0.5-1.0 mg selenium per week fully prevented this disorder, selenium appears to be required at about 60-120μg per day. Selenium uptake from food largely depends on the chemical form of selenium present (Yu *et al.,* 1988).

1.7 Mouse 56 kDa selenium binding protein and acetaminophen-binding protein

Selenium proteins can be divided into two different types. One type is called selenium containing protein, such as selenoprotein P and glutathione peroxidase. The characteristics of this type of selenium protein is that the selenium is present as selenocysteine in the protein (Fig1-1 shows the selenocysteine structure). An in-frame TGA codon represents the amino acid selenocysteine. The most important of the nucleotide sequence was the presence of some in-frame TGA codons in this group of selenium protein. Another type of selenium protein is called selenium binding proteins such as mouse 56 kDa selenium binding protein that are tightly bound with selenium in the body. The characteristics of this group of selenium binding protein is the absence of TGA codons in the nucleotide sequence. The study in selenium containing protein such as selenoprotein P is very advanced but research in the group of selenium binding protein, both in its function and metabolism are not well established (Lanfear *et al.*, 1993).

Fig 1-1 The structure of selenocysteine residue



A review on the mouse 56 kDa selenium binding protein for my reseach project is very useful in understanding their putative role of the human homolog, because mSP56 is similar to hSP56 in nucleotide sequence. From the previous study we know that the cDNA clone coding for the mouse 56 kDa selenium binding protein was isolated by screening a λZap mouse liver library (Bansal *et al*, 1990). Their cloned cDNA represented the complete message RNA. The correct reading frame was verified by alignment of the deduced amino acid sequence with that of peptides sequenced from the purified protein. The DNA sequence for mouse 56 kDa selenium binding protein did not contain an in-frame TGA codon that would code for selenocysteine, as occurs in the prototypic selenoprotein, glutathione peroxidase. Hydropathy analysis suggested that the protein was not a membrane-spanning protein. Using antibodies specific for mSP56, it was evident that the mSP56 was not induced by selenium and was not dependent upon dietary selenium levels (Lanfear *et al.,* 1993).

A full-length cDNA encoding a 56 kDa liver protein implicated in the detoxification of acetaminophen (AP56) has been cloned by virtue of its similarity to the 56 kDa selenium-binding protein (mSP56): in fact, the deduced AP56 amino acid sequence differs at only 14 residues from mSP56. Isolation of genomic DNA recombinants from a Balb/c mouse cosmid genomic DNA library shows that mSP56 and AP56 are encoded by two different genes. Using reverse transcription/PCR with oligonucleotide primers that distinguish the AP56 and mSP56 mRNAs shows that the mSP56 mRNA is highly expressed in liver, kidney and lung; whereas the AP56 mRNA is mainly expressed in liver. Both mRNAs tend to be down regulated in liver cell lines but remain high in DEN-induced liver tumours *in vivo.* The relevance of these findings is evaluated in terms of the postulated functions of the two proteins in mediating the anti-carcinogenic effects of selenium and detoxification mechanisms (Lanfear *et al.*, 1993).

Chapter 2

Materials and methods

2.1 Plating out the cDNA library

2.1.1 Media, buffers and solutions

♦ 1X TBE buffer

For 1 litre of 1X TBE buffer, 10.8g Tris base (Sigma), 5.5g boric acid (Riedel-de Haen) and 2 ml 0.5M EDTA (pH8.0) (Sigma) was made up to 1 litre with deionized water.

♦ Ampicillin (Sigma)

A stock of 25mg/ml of ampicillin was prepared by dissolving 2g ampicillin with 80 ml deionized water. The solution was stored at -20°C.

♦ X-gal (Promega)

It was purchased from Promega of 50 mg/ml. Then it was stored at -20°C.

♦ IPTG (Gibco BRL)

For plating of the library, a stock of 20 mg/ml was made. IPTG was dissolved in deionized water and then was sterilized by passing through 0.2μm filter membrane (Millipore). The solution of IPTG was stored in -20°C.

♦ **Suspension medium (SM)**

For 1 litre of SM solution, 5.8g of sodium chloride (Riedel-de Haen), 2g of magnesium sulphate (Riedel-de Haen), 50 ml of 1M Tris-HCl (Sigma) and 0.1g gelatin (Sigma) was made up to 1 litre with deionized water.

♦ **LB medium and LB plates**

For 1 litre of LB medium, 10g of bacto-tryptone (Difco), 5g of yeast extract (Difco) and 10g of sodium chloride (Riedel-de Haen) was made to 1 litre with deionized water. For LB plates, 15g of agar (Sigma) was added. The medium was made up to 1 litre with deionized water and then was autoclaved. The autoclaved LB medium with agar was equlibrated at 55°C. The medium was then poured onto Petri dishes inside the hood.

2.1.2 Bacteriophage clones preparation

The plating method basically followed that of Sambrook *et al.* (1989). Bacterial host *E.coli* Y1090 *hsdR* was grown overnight at 37°C, 280 rpm in 2 ml LB supplemented with maltose (Sigma) (0.2%) and magnesium sulphate (10mM) (Riedel-de haen). The overnight culture was then diluted

three fold by LB with 0.2% maltose and 10 mM magnesium sulphate. It was then incubated at 37°C, 280 rpm for another two hours. Then the agar plates to be poured were pre-warmed at 37°C. LB top agarose was also melted in a microwave oven and an aliquot of 3 ml top agarose was equilibrated in 55°C. After two hours of incubation, 100 pfu of the fetal human heart cDNA library was added to the culture and mixed gently. The culture was then incubated at 37°C without shaking for 20 min. Upon pouring of the library, IPTG (final concentration: 70μg /ml), X-gal (final concentration: 300 μg /ml) and ampicillin (final concentration: 50 μl /ml) were added to the equilibrated molten top agarose. The culture and the top agarose were then mixed and poured into the pre-warmed LB agar plates. The plates were allowed to cool down for 5 minutes and incubated at an inverted position at 37°C overnight. The colourless plaques were picked with sterile Pasteur pipettes on the next day. The plaques were resuspended in 50 μl of SM and were stored at 4°C.

2.2 cDNA clone amplification by PCR

The inserts of each cDNA clone were amplified using a pair of primers which is complementary to flanking sequences of the *NotI* and *SalI* cloning site of λgt 22A vector. The sequence of the primers are:

Forward: 5' -ATTGGTGGCGACGACTCCTGGA- 3'
Reverse: 5' -TTTGACACCAGACCAACTGGTA- 3'

PCR was done by using Taq DNA polymerase (Pharmacia) and the composition of reaction mix was:

Deionized water	36μl
10X reaction buffer * (Pharmacia)	5μl
2 mM dNTP mix (Pharmacia)	5μl
2.5 μM Primer mix (forward & reverse)	2μl
Phage stock	2μl
Taq DNA polymerase (Pharamacia)	(2U) 0.4μl

* 500mM KCL, 15mM $MgCl_2$, 100mM Tris-HCl (pH 9.0 at room temperature)

The reaction mix was then overlaid with 2 drops of mineral oil (Sigma).

The temperature profile of PCR was:

Step 1: 94°C for 5 min.
Step 2: 95°C for 36 sec.
Step 3: 50°C for 36 sec.
Step 4: 72°C for 1 min. 30 sec.

Step 2 to step 4 were repeated for 35 times and then followed with a final extension of 72°C for 10 min. The success of PCR was checked by running the PCR products in 1% agarose gel stained with ethidium bromide and checked with UV. Clones with one bright single band were chosen to perform cycle sequencing reaction.

2.3 Cycle sequencing of PCR products

2.3.1 Media, buffers and solutions

♦ **DNA size markers**

DNA size marker λ/HindIII-∅X174/HaeIII marker was purchased from stratagene (Cat#.201107) and λ/HindIII-pUC18/Sau3AI marker.

♦ **40 % acrylamide solution**

95g of acrylamide (Pharmacia) and 5g of bis-acrylamide (Pharmacia) were dissolved in deionized water and made up to 250 ml. It was stored at 4°C.

♦ **6 % urea-acrylamide solution**

The solution was made by dissolving 108g urea (Pharmacia) into 60 ml 40 % acrylamide solution, 24 ml of 10X TBE solution and then made up to 400 ml with deionized water. It was then filtered by 0.2μm filter (Millipore).

2.3.2 Preparation of sequencing reaction

The cycle sequencing of PCR products was done by using one of the following fluorescent primers. The sequence of the primers were as follows:

Forward: 5' -GGTGGCGACGACTCCTGGAGCC- 3'

Reverse: 5' -GACACCAGACCAACTGGTAATG- 3'

The cycle sequencing was done by using cycle sequencing kits (Pharmcia Biotech) and the composition of reaction mix was as follows:

Deionized water	**10μl**
10x reaction buffer (Pharmcia)	**2μl**
4mM Sequencing primer	**2μl**
PCR product	**3μl**
2mM dNTP mix (Pharmcia)	**5μl**
DMSO	**2μl**
Taq DNA polymerase (Pharmcia)	**0.4μl**

6μl of the reaction mix was pipetted to separate Eppendorff tube which had 3μl of one of termination mix* (ddATP, ddCTP, ddGTP, ddTTP). They were then overlaid with one drop of mineral oil. 2mM either of the four ddNTP plus 100 M each dATP, dCTP, dGTP, and dTTP.

The profile of the cycle sequencing reaction was as follows:

Step1: 95°C for 36 sec.;

Step2: 50°C for 36 sec.;

Step3: 72°C for 90 sec.

Step1 to step 3 were repeated 25 times and was then followed with a final extension of 72°C for 10 min. After the reaction ended, 5μl of stop solution was added.

2.4 Gel electrophoresis in the automated A.L.F. sequencer

6% urea-polyacrylamide gel was made by mixing 60 ml 6% urea-acrylamide solution, 280µl 10% APS and 70µl TEMED. The running buffer used was 0.6X TBE. The sequencing gel was casted and set for 60 minutes. Then the clamps for the comb were removed and the glass plate around the comb was cleaned. The upper reservoir was attached. After placing the lower reservoir into the electrophoresis unit and attaching the anode wire, the 0.6X TBE buffer was poured into the lower reservoir. The gel cassette was fitted into the electrophoresis unit and the two vertical etched lines were aligned. Afterward, the thermoplate was connected and the upper electrode was attached to the cathode connector. The main power switch was switched on and the computer was turned on. The laser beam was adjusted and the 0.6X TBE buffer was poured into the upper reservoir. Water was preheated and the comb was removed. The walls were examined and rinsed. The samples were denatured by heating at 95°C for 3 min and chilled on ice before loading. Then 8µl of each sample was loaded into the wells in the order A, C, G, T. The sequencing gels were

run at 40°C and 34 W for six hours. The laser power was set to 3 mW. After six hours, the raw data was processed (Smith L. M. 1986). The sequences which are at least 150 bp with an accuracy greater than 97% were marked and exported to a floppy disk.

2.5 DNA sequence analysis

The raw cDNA sequences of each clone were edited with the software A.L.F. manager and the edited sequences were sent for comparison with the GenBank and EMBL nucleotide and protein databases using the NCBI BLAST electronic mail servers (Altschul *et al.*, 1990). The address of the BLAST server is "blast@ncbi.nlm.gov". The sequences reported were searched against the nonredundant nucleotide sequence database by the BLASTN program with an EXPECT value of 0.75. The EXPECT value is the statistical significance threshold for reporting matches against database sequence. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Sequence similarity identified by the BLAST programs was considered statistically significant with a Poisson P-value smaller than 0.01. The Poisson P-value is the possibility of a score occurring by chance, given the number of residues in the query sequence and the database (Adams *et al.*, 1991). The NCBI RETRIEVE E-mail server (retrieve@ncbi.nlm.nih.gov) was used to retrieve the information of a particular match. The information retrieved can help us to determine whether the fetal human cDNA clones are full length or not.

2.7 Transformation of plasmid into competent *Escherichia coli*

Frozen competent cells were removed from a deep freezer and thawed at room temperature until the cell suspension was just thawed. The tube was then placed on ice. DNA solution with volume less than 20μl was added. After mixing, the tube was chilled on ice for 45 minutes. The tube of cells were heat shocked at 42°C for exactly 90 seconds and then chilled on ice for 2 minutes. The contents of the tube were mixed with 800μl LBG medium (0.2ml 1M glucose/10ml LB). The tube was then incubated at 37°C with shaking (250 r.p.m.) for 45 minutes. The transformed cells (diluted to a suitable concentration) were spread onto LB agar plates containing the appropriate antibiotics.

2.8 Mini-preparation of plasmid DNA (Sambrook *et al.*,1989)

A single bacterial colony was transferred from the agar plate into 3 ml of LB medium containing the appropriate antibiotic. The culture was incubated overnight at 37°C with shaking (250 r.p.m.). 1.5ml of the culture was transferred into a microfuge tube and then centrifuged at 4000 x g for 30 seconds. The medium was removed by aspiration. The pellet was suspended in 100µl ice-cold Solution I (50mM glucose, 25mM Tris-HCl pH8.0, 10mM EDTA pH8.0) by vigorous vortexing. The tube was left on ice for 5 to 10 minutes. 200µl freshly prepared Solution II (0.2M NaOH, 1% SDS) was added. The contents were mixed by inverting the tube rapidly five times and then left on ice for 5 minutes. 150µl ice-cold Solution III (60 ml 5M potassium acetate, 11.5 ml glacial acetic acid and 28.5 ml water in each 100 ml) was added. The tube was vortexed in an inverted position for 10 seconds. The lysate was stored on ice for 15 minutes. Bacterial cell debris was spun down by centrifugation in a microfuge at 4000 x g for 5 minutes at 4°C. The clear supernatant was transferred into another microfuge tube

into which 2 volumes of ice-cold absolute ethanol was added to precipitate the DNA. The contents of the tube was mixed gently and then stored at minus 20°C for 20 minutes. Precipitated DNA was pelleted by centrifugation in a microfuge at 4000 x g for 5 minutes at 4°C. Supernatant was discarded and the pellet was washed with 1 ml of ice-cold 70% ethanol. The pellet was vacuum dried for 20 minutes. The DNA powder was dissolved in a minimum volume of water containing RNase (20μg/ml).

2.9 Large scale plasmid DNA preparation by QIAGEN™

A bacterial colony harbouring the plasmid of interest was transferred to 250 ml LB medium containing a suitable antibiotic, for example, ampicillin (50 mg/ml). The culture was incubated at 37°C overnight with shaking (250 r.p.m.). The bacteria suspension was spun down by centrifugation at 3500 x g. for 20 minutes at 4°C. The bacteria pellet was resuspended in 4.0 ml ice-cold P1 buffer (100μg/ml Rnase A in 50 mM Tris-HCl, 10 mM EDTA, pH8.0). vigorous vortexing was essential. 4.0 ml P2 buffer (1% SDS, 200mM NaOH) was added afterwards. The contents were mixed by inverting the tube, which were then allowed to stand at room temperature for 5 minutes. 4.0 ml chilled P3 buffer (3.0M potassium acetate, pH5.5) was added to the tube. The contents were mixed by immediate inversion, followed by incubation on ice for 15 minutes. The bacteria lysate was centrifuged at 30,000 x g for 30 minutes at 4°C.

A QIAGEN-tip 100 (Qiagen Inc, Dusseldorf, FRG) was equilibrated with 3.0 ml QBT buffer (750mM NaCl, 50mM MOPS, 15% ethanol, pH7.0, 0.15% Triton X-100). The clear bacterial lysate was applied onto the column. The column

was washed with 10 ml QC buffer (1.0M NaCl, 50mM MOPS, 15% ethanol, pH7.0) twice. DNA was eluted out of the column by 5.0 ml QF buffer (1.25M NaCl, 50mM Tris, 15% ethanol, pH 8.2). The DNA solution was mixed with 0.7 volume of isopropanol which had been equilibrated to room temperature. The DNA/isopropanol mixture was spun at 20000 x g for 30 minutes at 4°C. The DNA pellet was washed with ice-cold 70% ethanol and stored at 4°C.

2.10 Cloning the human 56 kDa selenium binding protein (hSP56) into the pAED4 vector

2.10.1 Bacterial strains and vectors

♦ **Plasmid vector pAED4**

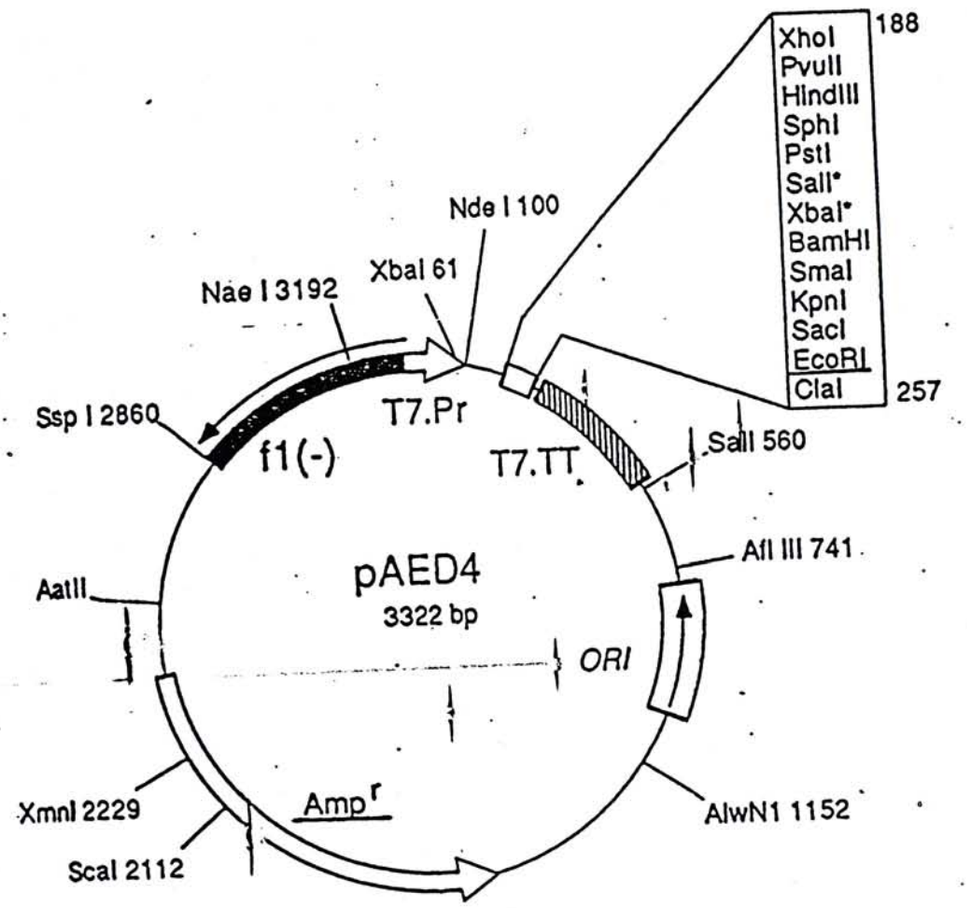
The vector is an expression vector which can be selected by ampicillin and induced by IPTG. Upon IPTG induction, the strong T7 promoter will recruit T7 RNA polymerase which allow the over-expression of the target protein. The vector is obtained from Don Doering and Paul Matsudaira. The genetic map of pAED4 is shown on Fig. 2-1.

♦ ***Escherichia coli*** **JM109** ***E. coli*** **JM109 is a common host for transformation. The genotype of** ***E. coli*** **JM109 is** ***rec*****AI** ***supE*****44** ***end*****AI** ***hsdR*****17** ***gyrA*****96** ***rel*****AI** ***thi*** **Δ (*lac-proAB*) F'{*traD*36 *proAB*$^+$ *lacI*q *lacZ*ΔM15}.**

♦ ***Escherichia coli*** **BL21(DE3)pLysS and BL21(DE3)pLysE**

pAED4 is used as the vector (Studier *et al.*, 1990) for gene expression in *E.coli*. It can be selected by chloramphenicol. T7 RNA polymerase and T7 lysozyme genes were incorporated into the bacterial host. Their genotypes are as follows:

Fig. 2-1. Schematics diagram fo the pAED4 vector. T7.Pr: T7 promotor; T7.TT: T7 transcriptional terminator; f1(-): intergenic region.



Strain	Genotype
BL21 (DE3)pLysS	F⁻, *ompT*, *hsdS* (*r*B⁻*m*B⁻), *gal dcm* (DE3),p*LysS* (Cm^r)
BL21 (DE3)pLysE	F⁻, *ompT*, *hsdS* (*r*B⁻*m*B⁻), *gal,dcm* (DE3), p*LysE* (Cm^r)

2.10.2 Media, buffers and solutions

♦ **Chloramphenicol (Sigma)**

A stock of 34 mg/ml of chloramphenicol was made by dissolving 340 mg of chloramphenicol with 10 ml ethanol. It was stored at -20°C.

♦ **Enzymes**

*Eco*RI (25U/μl) and *Nde*I (10U/μl) were purchased from Pharmacia. T4 DNA ligase (1U/μl) was purchased from Gibco.

♦ **LBA medium and LBA plates**

For LBA medium, ampicillin was added into the LB medium which was autoclaved. The final concentration of LBA is 50μg/ml. For LBA plates, agar with autoclaved LB medium was equilibrated at 55°C. Ampicillin of 50μl/ml was added when necessary. The solution was then poured onto Petri dishes inside the hood.

♦ **LBAC medium and LBAC plates**

For LBAC medium, ampicillin (200μg/ml) and chloramphenicol (34μl/ml) were added into the LB medium which was autoclaved. For LBAC plates, agar with autoclaved LB medium was equilibrated at 55°C. The ampicillin of 200μl/ml and chloramphenicol of 34 μg/ml were added. The solution was then poured onto Petri dishes inside the hood.

♦ **SOB Medium**

The composition of one litre of SOB is as follows:

Bacto-trytone (Difco)	**20g**
Yeast extract (Difco)	**5g**
Sodium chloride (Ridel-de Haen)	**0.5g**
250mM Potassium chloride (Ridel-de Haen)	**20 ml**

It was made up to one litre with deionized water and autoclaved before use. After that, 5 ml of solution of 2M $MgCl_2$ (Riedel-de Haen) was added and mixed.

♦ **SOC Medium**

The composition of 1 litre of SOC is as follows: 20 ml of 1M glucose (Riedel-de Haen) after sterilization was added to 980ml of SOB solution after autoclave.

♦ **Solution I**

50mM glucose (Merck), 25mM Tris-HCl (Sigma) (pH8.0), 10mM EDTA (Sigma)(pH8.0). The solution was stored at 4°C.

♦ **Solution II**

0.2M NaOH (Merck), 1% SDS (Sigma).

♦ **Solution III**

It was made by 60 ml of 5M potassium acetate (Riedel-de Haen) and 11.5 ml of glacial acetic acid (Merck) into 28.5 ml of deionized water.

♦ **TFB solution**

10 mM MES (Sigma)(pH6.3), 45 mM $MnCl_2$(Riedel-de Haen), 10 mM $CaCl_2$(BDH), 100mM Kcl (Riedel-de Haen), 3mM hexamminecobalt chloride(Sigma).

♦ **DnD solution**

For 10 ml DnD solution, 1.53g DDT(Sigma), 9 ml DMSO(BDH) 100μl 1 M potassium acetate (pH7.5)(Ridel-de Haen) were made up to 10 ml with deionized water.

♦ **1XTAE buffer**

For 1 litre 1X TAE buffer , 4.84g Tris base (Sigma), 1.142g glacial acetic acid (Riedel-de Haen) and 2 ml of 0.5M EDTA (pH8.0)(Sigma) was made up to 1 litre with deionized water.

♦ **1.5M Tris-HCl (pH8.8)**

For 100 ml of 1.5M Tris-HCl solution, 18.15g Tris base (Sigma) was dissolved in 100 ml deionized water.

♦ 1.0M Tris-HCl (pH6.8)

For 100 ml of 1.0M Tris-HCl solution, 6g Tris base (Sigma) was dissolved in 100 ml deionized water.

♦ 1X Tris-glycine buffer

For 1 litre 1X Tris-glycine buffer, 3.02g Tris base (Sigma), 18.8g glycine (Sigma) and 10 ml 10 % SDS (Sigma) was made up to 1 litre with deionized water.

♦ 10 % (w/v) SDS

10g SDS (Sigma) was dissolved in deionized water and made up to 100ml.

♦ Protein gel staining solution

0.2 % (w/v) Commassie blue (Sigma), 25% (v/v) methanol (Merck), 10 % (v/v) acetic acid (Merck) was made in deionized water.

♦ Protein gel destaining solution

25 % (v/v) ethanol (Merck), 0.08 % (v/v) acetic acid (Merck) was made in deionized water.

♦ Lysis buffer

10 mM KCl (Riedel-de Haen), 10 mM potassium phosphate (Riedel-de Haen) (pH7.2), 10 mM DDT (Sigma).

2.10.3 Primer design and PCR

A primer which has its 3' end complementary to the 5' end of the open reading frame (ORF) of the cDNA of hSP56, an *Nde*I site overlapping the ATG start codon and a GC clamp at the 5' end of the primer was designed. The sequence of the cloning primer is as follows:

5' -TAGGGCCATATGGCTACGAAATGTGGGAATTG- 3'

*Nde*I site

At the other end, an oligo dT primer which acted as an universal primer to prime to the polyA tail of the mRNA was designed. An *Eco*RI site was present to enhance later subcloning step. The sequence of the dT primer was as follows:

5' -TAGGGCGAATTCTTTTTTTTTTTTTTTTTT- 3'

EcoRI site

The composition of the PCR reaction mix was as follows:

Deionized water		33.7μl
plate lysate of hSP56		5μl
10X reaction buffer*(Pharmacia)		5μl
5μM primer mix (cloning and dT primers)		5μl
2mM dNTP		5μl
Taq DNA polymerase	(2U)	0.3μl

*500mM KCl, 15mM $MgCl_2$, 100mM Tris-HCl (pH 9.0 at room temperature)

The reaction mix was then overlaid with 2 drops of mineral oil.

The temperature profile of the expression-cassette PCR was as follows:

Step 1:	94°C for 5min.;
Step 2:	95°C for 36 sec.;
Step3:	50°C for 36 sec.;
Step 4:	72°C for 1 min. 30 sec.

Step 2 to 4 were repeated for 35 times and then followed by a final extension of 72°C for 10 min. The success of PCR was checked by running the PCR products in a 1% agarose gel stained with ethidium bromide and was checked under UV.

2.10.4 Purification of PCR products by Geneclean

The method used basically followed the instructions of the manufacturer (BIO 101 Inc). PCR product of hSP56 was resolved in 1% agarose TAE gel. The gel block corresponding to hSP56 was cut and put into an Eppendorff tube. Three volumes of NaI solution was then added and the gel block was melted by incubation at 55°C for 5 min with agitation.

20μl glass milk was then added and kept at room temperature for 5 min. with agitation. PCR products should bind to glass milk. The glass milk was then pelleted by a brief spin and the supernatant was discarded. The pellet of glass milk was washed by resuspending it in 250μl NEW wash solution (provided by manufacturer) which contains NaCl, Tris, EDTA and ethanol and then a brief spin. The washing was repeated two more times. To elute the PCR product, 20μl of deionized water was used to resuspend the pellet of glass milk which was then incubated at 55°C for 5 min. The glass milk was pelleted by a brief spin and the supernatant was collected into another Eppendorff tube. The elution step was repeated once using 10μl of deionized water. The eluted PCR product was then pooled.

2.10.5 Restriction digestion of purified PCR product and pAED4

Both purified PCR product and pAED4 vector DNA were digested with *NdeI* and *Eco*RI (S.K.W. Tsui Ph.D. thesis, 1995). For the purified PCR product, the following reaction mix was prepared:

10X One-Phor-All Plus buffer*(Pharmacia)		4μl
Deionized water		6μl
Purified PCR product	(150ng)	9.0μl
*Eco*RI (Pharmacia)	(12.5U)	0.5μl
*Nde*I (Pharmacia)	(5.0 U)	0.5μl

For the pAED4 vector, the following reaction mix was prepared:

10X One-Phor-All Plus buffer*(Pharmacia)		4μl
Deionized water		5μl
pAED4 vector	(1μg)	10μl
*Eco*RI (Pharmacia)	(2.5U)	0.5μl
*Nde*I (Pharmacia)	(5.0 U)	0.5μl

*10X One-Phor-All Plus buffer: 100mM Tris-acetate pH7.5), 100 mM magnesium acetate, 500mM potassium acetate.

The reaction mixes were incubated at 37°C for 3 hours. The restricted PCR product and pAED4 DNA were than purified by Geneclean II™ as in section 2.10.4.

2.10.6 Ligation and transformation of hSP56

There were four ligation reactions: D1, double-digested pAED4 without ligase; D2, double-digested pAED4 with ligase; D3, EcoRI-digested pAED4 with ligase; D4, double-digested pAED4 and PCR product with ligase. For D4, the ratio of vector and the insert was set to 1:3. The four reaction mixes were as follows: (unit: μl).

	D1	D2	D3	D4
Deionized water	-	18	18	18
Vector	5	5	5	5
5X ligase buffer*	4	4	4	6
Insert	-	-	-	16
T4 DNA ligase (Gibco) (1U/μl)	-	2	2	2

*5X buffer: 250mM Tris-HCl (pH7.6), 50mM $MgCl_2$,5mM ATP, 5mM DDT, 25%(w/v)PEG-8000.

The reaction mixes were then incubated at 16°C overnight. The entire reaction mixes were used for the transformation experiments.

The preparation of competent cells and the transformation method were essentially the same as that of Sambrook *et al.* (1989). Firstly, single colony of *E. coli* JM 109 was inoculated into 1 ml of SOB. It was incubated at 37°C overnight. Then 0.1ml of overnight culture was inoculated into 10ml SOB and was incubated at 37°C for 3 hours. The bacterial cells were collected by centrifuging at 5000 r.p.m. in a 20-2 rotor (Hitachi) for 5 minutes. at 4°C. The pellet was resuspended in 5 ml cold TFB and was incubated on ice for 10 minutes. The bacterial cells were then collected by centrifuging at 5000 r.p.m. in a 20-2 rotor (Hitachi) for 5 minutes. at 4°C. The pellet was then

resuspended in 0.6ml TFB. 21μl of DnD solution was added and mixed. The bacterial cells were incubated on ice for 10 minutes. Another 21μl of DnD solution was added to the cells and mixed again, and the tube was then incubated on ice for 20 minutes. Aliquots of 210μl of the bacterial cell were pipetted into 1.5ml Eppendorff tubes. The four ligation reaction mixes were then pipetted into separate Eppendorff tubes of bacterial cell and mixed. The reaction mixes were incubated on ice for 1 hour and then heat-shocked at 42°C for 2 minutes. It was kept on ice for 2 minutes. Then 800μl of SOC solution was added into each Eppendorff tube and incubated at 37°C for 45 minutes. Lastly, 1μl, 10μl, 100μl were spread on LBA plates and incubated at 37°C overnight. The successfully tranformed bacteria were grown in 2 ml LBA medium at 37°C overnight and the plasmids were prepared by minipreparation (Sambrook *et al.*, 1989).

2.10.7 Screening and purfication of pAED4-hSP56

Mini-preparation (Sambrook *et al.*, 1989) was used to screen for recombinant pAED4-hSP56. 1 ml of overnight culture was centrifuged at 3500 x g at 4°C for 30 sec. using a

microcentrifuge. The pellet was then resuspended in 100μl solution I and mixed well. 200μl of solution II was added and mixed by inverting the Eppendorff tube several times. Then 150μl solution III was added and mixed. The solution was incubated on ice for 5 minutes. It was then centrifuged at 20000 x g at 4°C for 10 minutes. The supernatant was collected and the plasmid was precipitated by adding 2 volumes of ice-cold absolute ethanol and stored on ice for 10 minutes. The solution was then centrifuged at 20000 x g at 4°C for 10 minutes. The pellet was washed with 70% ethanol and dried. Lastly, the pellet was dissoved in the desired volume of deionized water. The plasmid was stored in -20°C.

The mini-preparation of pAED4-hSP56 was performed by using a commercially available plasmid-adsorbing column (Qiagen). 30 ml of overnight culture in LBA bearing the desired plasmid was inoculated into 500 ml LBA. It was incubated at 37°C for 12 hours. The culture was then centrifuged at 2500 x g at 4°C for 10 minutes. The pellet was resuspended in 100 ml ice-cold 1X STE and then centrifuged at 2500 x g at 4°C for 10 minutes. The pellet was then

resuspended in 4 ml P1 solution (100μg/ml Rnase A [Sigma], 50mM Tris-HCl, 10mMEDTA [pH8.0]) and mixed. Then 4 ml of P2 solution (200mMNaOH,1% SDS) was added and mixed. The solution was kept at room temperature for 5 minutes. 4 ml of P3 solution (2.55 M potassium acetate [pH4.8]) was added and mixed gently. The solution was then centrifuged at 35000 x g at 4°C for 30 minutes. The suspernatant was saved and loaded onto a QIAGEN-pack 100 cartridge which was pre-equilibrated with QB solution (750mM NaCl, 50mM MOPS, 15% ethanol [pH7.0]). The flow rate was kept at 1 drop/sec. The cartridge was washed two times with 4 ml of QC solution (1.0M NaCl, 50mM MOPS, 15% ethanol [pH7.0]). The flow rate was kept at 2 drops/sec. Then the plasmid was eluted from the cartridge by using 2 ml of QF solution (1.2M NaCl, 50 mM MOPS, 15% ethanol [pH8.0]). The flow rate was kept at 1 drop/sec. 0.5 volume of isopropanol was added to the elute and kept on ice for 30 minutes. The solution was centrifuged at 35000 x g in 20-2 rotor (Hitachi) at 4°C for 30 minutes. The pellet was washed with 70% ethanol and dried. The pellet was dissolved in an appropriate volume of deionized water and stored at –20°C.

2.11 Expression of hSP56

2.11.1 Induction of hSP56 expression

The expression system that we have utilized was the T7 RNA polymerase expression system which can overexpress the target protein using *E coli* as the host (Studier *et al.*,1990). The procedure of hSP56 expression was basically the same as that of Studier *et al.* (1990). Firstly, single isolated transformed clones of BL21(DE3)pLysS and BL21(DE3) pLysE plated in LBAC plates were picked and inoculated into 5ml of LBAC medium which were allowed to grow at 37°C at 250 r.p.m. overnight. The overnight cultures were inoculated into another flasks of 50 ml LBAC medium at a concentration of 1% culture. The cultures were allowed to grow at 37°C at 250 r.p.m. until OD_{600} reached 0.5-1. It took approximately 3 hours. 50 ml culture was then inoculated into 1 litre of LBAC medium and allowed to grow at 37°C at 250 r.p.m. until OD_{600} reached 0.5-1. It took approximately 5 hours. A final concentration of 0.4 mM IPTG and 0.2mM $ZnSO_4$ were added to start the induction. The induction lasted three hours. 1 ml of culture was pelleted and resuspended into 100μl of 1X SDS gel-loading buffer to check if hSP56 was expressed. The remaining cultures were

collected by centrifuging the culture at 3500 x g at 4°C for 10 minutes. The pellets were then resuspended in 1/50 of volume of lysis buffer. The bacterial cells were sonicated and the cell debris was removed by centrifugation at 20000 x g for 30 minutes at 4°C.

2.11.2 SDS-PAGE and protein detection

Laemmli buffer system (1970) was used for the analysis of proteins. The composition of the 15 % resolving gel is as follows:

Deionized water	3.55 ml
1.5M Tris-HCl (pH8.8)	2.5 ml
10 % (w/v) SDS	0.1 ml
40% acrylamide/bis- acrylamide solution	3.75 ml
10% (w/v) ammonium persulphate (Serva)	0.1 ml
TEMED (Pharmacia)	4.0µl

The mini-gel was casted as directed from Bio-Rad. The resolving gel was poured between the mini-gel plates and was layered with water-saturated isobutanol (Mallinckrodt). The gel was set for 45 minutes. Afterwards, the layer of water-saturated isobutanol was cleaned up and the plates were dried. 5 % Stacking solution is as follows:

Deionized water	3.64 ml
1.0 M Tris-HCl (pH6.8)	0.63 ml
10 % (w/v) SDS	0.05 ml
40 % acrylamide/bis-acrylamide solution	0.61 ml
10 % (w/v) ammonium persulphate (Serva)	0.05 ml
TEMED (Pharmacia)	5.0µl

It was then poured between the mini-gel plates and a comb was inserted from the top of the gel. The gel was set for I hour. Just before running of protein samples, the samples were boiled for 5 minutes. The protein gels were then run in 1X Tris-glycine buffer of constant current 70 mA (for two gels) for about 1 hour.

The gel may be stained with Coomassie blue. Coomassie blue staining method can detect proteins present in the polyacrylamide gels at a μg level. Coomassie blue staining was done by immersing the protein gels into the protein gel staining solution overnight with gentle agitation. The protein gels were then destained in protein gel destaining solution with gentle agitation. The protein destaining solution was changed several times until the protein gels were destained to the desired intensity.

2.12 Northern hybriddization of hSP56

2.12.1 Animal & human tissue

Spraque-Dawley rats were used to prepare total RNA. The fetal human heart tissue was from an aborted fetus.

2.12.2 Preparation of medias, buffers and solutions

For those solutions which were involved in the analysis of RNA, they were treated with DEPC (Sambrook *et al.,* 1989). Briefly, DEPC (Fluka) was added to solutions (0.1% (v/v). The solutions were agitated and autoclaved. However, those solutions which were composed of Tris base were not treated in this way. Instead, Tris base was directly dissolved in DEPC-treated deionized water.

- 0.5 M EDTA (Sigma)

186.1 g EDTA disodium salt/ dihydrate (Sigma) was dissolved in deionized water and make up to 1 litre (pH 8.0). It was then autoclaved.

- 5X formaldehyde gel-running buffer

0.1 M MOPS (pH7.0) (Sigma), 40 mM sodium acetate (Merck), 5mM EDTA (pH8.0) (Sigma).

♦ **Formaldehyde gel-loading buffer**

50 % glycerol (Sigma), 1 mM EDTA (pH8.0) (Sigma), 0.25 % bromophenol blue (Sigma), 0.25 % xylene cyanol FF (BDH).

♦ **20X SSC**

For 1 litre 20X SSC, 175.3g sodium chloride (Riedel-de Haen) and 88.2g sodium citrate (Amresco) were dissolved in 800 ml deionized water. They were then adjusted to pH 7.0 with 10M NaOH (Merck). The volume was adjusted to 1 litre with deionized water and was autoclaved.

♦ **20X SSPE**

For 1 litre 20X SSPE, 175.3g sodium chloride (Riedel-de Haen), 27.6g NaH_2PO_4-$2H_2O$ (Riedel-de Haen) and 7.4g EDTA (Sigma) were dissolved in 800 ml deionized water. The pH was then adjusted to 7.4 with 10M NaOH (Merck). The volume was adjusted to 1 litre with deionized water and was autoclaved.

♦ **1X STE**

0.1M NaCl (Riedel-de Haen), 10mM Tris-Cl (pH8.0), 1mM EDTA (pH8.0) (Sigma).

♦ **1X TE**

10mM Tris-Cl (Sigma), 1mM EDTA (pH8.0) (Sigma).

♦ **CsCl/EDTA solution**

5.7MCsCl (Gibco BRL), 0.01M EDTA (pH7.5) (Sigma)

♦ **RNA extraction buffer**

To prepare 50 ml, 25g of guanidinium thiocyanate (Gibco BRL), 2.64 ml of 10 % sodium lauryl sarcosinate (Sigma), 0.36 ml of β-mercaptoethanol (Riedel-de haen), 1.76 ml of 0.75M sodium citrate (Amresco) was added.

♦ **Guanidinium thiocyanate homogenization buffer**

4.0M guanidinium thiocyanate (Gibco BRL), 0.1M Tris-Cl (pH7.5)(Sigma), 1% β-mercaptoethanol (Riedel-de Haen).

♦ **100X Denhardt's reagent**

For 500 ml of 100X Denhardt's reagent, 8g of Ficoll (Sigma), 8g of polyvinylpyrrolidone (Sigma) and 8g of bovine serum albumin (Sigma) was dissolved in 500 ml DEPC-treated water. It was then sterilized by filtration and stored at -20°C.

♦ **Solution A**

0.32 M sucrose (Sigma), 10mM Tris-HCl (Sigma) (pH7.5), 5mM $MgCl_2$ (Riedel-de Haen), 1% TritonX-100(v/v) (Sigma).

♦ **Solution B**

4Mguanidinium thiocyanate(Gibco BRL), 25mM sodium acetate(pH4.8),0.84%(v/v) β-mercaptoethanol (Riedel-de haen).

2.12.3 Preparation of total RNA

The AGPC method (Chomczynski and Sacchi, 1987) was used to prepare total RNA from rat tissues. This method is suitable for small scale preparation of total RNA . For 100mg of tissue, 1 ml RNA extraction buffer was added for homogenization with a glass homogenizer. 100µl of 2M sodium acetate (pH4) (Merck) was added and mixed. 1 ml of Tris-saturated phenol (Gibco BRL) and 200µl of chloroform (Merck): isoamy alcohol (Merck) (49:1) was then added and the solution was vortexed for 20 sec. The solution was kept on ice for 15 minutes. The solution was then centrifuged at 15000 r.p.m. for 20 minutes at 4°C in a microcentrifuge. The upper aqueous layer was carefully transferred to another Eppendorff tube. 1 volume of ice-cold isopropanol (Riedel-de Haen) was added and the solution was kept at -20°C for 20 minutes. The solution was centrifuged at 12000 x g in a microcentrifuge for 10 minutes at 4°C. The supernatant was discarded and the RNA pellet was washed with 70 % ethanol (Merck). The pellet was then dried and dissolved in 0.5 % SDS or DEPC-treated water.

2.12.4 Formaldehyde agarose gel electrophoresis

The method used was based on that of Sambrook et al. (1989). The 1.5 % agarose/2.2 M formaldehyde (Riedel-de haen) gel was prepared by mixing 0.74g agarose (Sigma) with 30.6 ml deionized water and 9.9 ml 5X formaldehyde gel-running buffer. The agarose/water solution was melted by microwave for 2 minutes. at medium power. 9 ml formaldehyde was then added to the melted agarose solution and was poured to the gel cast. The gels were set in a flow hood. The RNA samples were prepared as follows:

RNA (40μg)	5.0μl
5X formaldehyde gel-running buffer	2.0μl
formaldehyde (Riedel-de Haen)	3.5μl
formamide (Amresco)	10.0μl

The sample were incubated at 65°C for 15 minutes. and then chilled on ice. 2μl of formaldehyde gel-loading buffer and 1μl of ethidium bromide (1mg/ml) was then added. The samples were run in the gel submerged in 1X formaldehyde gel-running buffer at 30V. The gel was collected when the dye front has migrated to 3/4 of the gel. The gel was inspected above the UV light box to see if the RNA was intact and well separated.

2.12.5 Preparation of radioactive probe

The method of random priming (Vogelstein and Feinberg, 1983) was used to prepare radioactive probes. The procedure followed the instruction of the manufacturer (Megaprime™, Amersham). Briefly, 2μl of template DNA obtained by using PCR primers of the following sequences 5'TAGGGCCATAGCCCAAGTGTCCCAAGTGC 3' and 5' AGGATGGGTCTCCACCACCTG 3' and the PCR product (about 25ng) was mixed with 5μl of 9-mer primer mix provided by the manufacturer and 26μl deionized water. The mixture was then boiled for 5 minutes. It was chilled on ice and then 10μl of labelling buffer (Amersham), 5μl of [α-P^{32}] dCTP (10 uCi. Amersham) and 2μl of Klenow fragment (Amersham) were added to the mixture. The reaction mix was incubated at 37°C for 1 hour. The reaction was stopped by adding 5μl of 0.2M EDTA. The reaction mix was then passed through a 1X STE equilibrated NucTrap© push column (Stratagene) to remove unincorporated nucleotides from the radiolabeled DNA probe. Then 75μl 1XSTE was added to the column to facilitate the collection of purified radioactive probe. The purified radiolabeled probe was then denatured by boiling for

3 minutes. It was then added to the hybridization solution after RNA transfer and prehybridization of the membrane. Although the exact CPM of the probe was not determined by Scintillation counting, a reading of 150-200 CPS (counts per second) on the hand held gamma radiation counter (mini monitor 900, Mini instruments LTD) is usually detectable when it is held 1 cm away from the labelled probe which is resuspend in 50 μl solution in a 200 μl (Sarstedt) microtube.

2.12.6 RNA transfer and Northern hybridization

The formaldehyde/agarose gel after electrophoresis of total RNA was rinsed for three times with DEPC-treated water. The gel was then immersed in 20X SSC for 20 minutes with gentle agitation. A 3MM Whatman paper was used to mark a bridge which transfer the buffer to the gel. The gel was placed on the bridge in an upside-down orientation. The transfer buffer used was 20X SSC. 3MM Whatman papers were soaked in 2X SSC for 3 minutes. Nylon membrane was equilibrated in 20X SSC for 10 minutes. It was carefully layered on the gel. Three sheets of 3MM Whatman paper pre-soaked in 2X SSC was layered on the Nylon membrane.Bubbles were carefully removed. Then a stack of

paper towel was added to the top of the 3MM Whatman papers. The transfer proceeded overnight. The efficiency of RNA transfer was checked by studying the intensity of fluorescence of EtBr that remained in the gel. The Hybond-N nylon membrane onto which the electrophoresized RNA were transferred was washed with DEPC-treated 6X SSC for 5 minutes. It was allowed to dry at room temperature for 15 minutes. The nylon menbrane was baked under vacuum at 80°C for 1 hour. The nylon membrane may be stained by first immersing the membrane in 5% acetic acid (Riedel-de Haen) for 10 minutes. Then the membrane was stained by membrane-staining solution (0.5 M sodium acetate {pH5.2} {Merck}, 0.04% methylene blue) for 5 minutes. The membrane was then destained in deionized water until desired results were obtained. Upon prehybridization, the nylon membrane was pre-wetted in 6X SSC for 15 minutes. It was then incubated at 42°C for 6-8 hours with the following prehybridization solution:

formamide (Amresco)	10 ml
100X Denhardt's reagent	1 ml
20X SSPE	6 ml
10% SDS (Sigma)	1 ml
10mg/ml sperm DNA* (Sigma)	0.2 ml
DEPC-treated water	1.8 ml

*10mg/ml sperm DNA was boiled for 5 minutes just before use.

After prehybridization, the membrane was hybridized with radio-labelled probe as prepared in section 2.12.5. The hybridization solution is described as follows:

formamide	10 ml
20X SSPE	6 ml
10% SDS	1 ml
DEPC-treated water	2.8 ml
radioactive probe	100μl

The hybridization was carried out at 42°C for 20 hours with gentle agitation. The nylon membrane was then washed 2 times with 1X SSC/0.1% SDS at room temperature for 5 minutes. It was then washed once with 0.1X SSC/0.1% SDS at room temperature for 5 minutes. Then the membrane was put into the cassette and autoradiographed at -70°C for 48 hours. After that, the cassette was removed to room temperature for about 2 hours before developing the film in the dark room.

2.13 Chromosomal mapping of the hSP56 gene

The chromosomal mapping of the hSP56 gene was performed by Dr.H.H.Q. Heng of SeeDNA Biotech Inc. (Ontario, Canada). The pAED4-hSP56 plasmid was biotinylated with dATP using the BRL (Gaithersburg, MD) BioNick labelling kit. The procedure for fluorescent in situ hybridization (FISH) detection was performed essentially as described by Heng *et al.* [1992] and Heng and Tsui [1993].

CHAPTER 3

RESULTS

3.1 The sequencing results of 553 cDNA clones

I have sequenced 553 adult human heart cDNA clones from the cDNA library using the automatic sequencer, assigned putative identities of all the clones, organized the database of these ESTs and analysed the results of the computer search. Novel sequences (no significant match with known gene) and sequences that match to other human ESTs only were submitted to GenBank. The criteria for submission were that sequences with no ambiguities for at least 150 base in length. A summary of results is shown in Table 3-1.

Table 3-1. Categories of cDNA from adult human heart cDNA library.

BLASTN search result	Number of clones	%
Alu repeat	16	2.9
Mitochondrial	161	29.1
Novel	153	27.7
Repetitive	9	1.6
Ribosomal RNA	7	1.3
Known	207	37.4
Total	**553**	**100.0**

3.2 Catalogues of genes expressed

I have analysed 346 novel genes and 207 known genes in total. To study the pattern of gene expression in the adult human heart, 207 clones are categorized into ten groups: contractile elements cytoskeleton and related proteins, extracellular matrix proteins, proteins involved in energy metabolism, hormones and proteins involved in hormonal regulation, proteins that participate in signal transduction and cell regulation, proteins of transcriptional and translational machinery, membrane associated proteins, proteins involved in other metabolism and miscellaneous proteins (Table 3-2). A summary of 207 clones homologous to known genes is shown in Table. 3-3.

Table 3-2. Expression profiles of adult human heart cDNA library

Category	Number of clones	(%)
Contractile element	35	16.9
Cytoskeleton related	18	8.7
Extracellular matrix	3	1.4
Energy metabolism	24	11.6
Hormones and hormonal control	10	4.8
Signal transduction/cell regulation	7	3.4
Transcription and translation	38	18.4
Membrane associated	11	5.3
Other metabolism	37	17.9
Miscellaneous	24	11.6
Total	**207**	**100.0**

Table 3-3. Human cardiac ESTs which are matched to known genes in the GenBank and EMBL databases. For the clone number, the first alphabet and the following three numbers indicate the identity of the clone in our library and the 'F' indicates the cDNA clones were sequenced by using forward primer.

Clone no	Gene-ID	Accession	Length(bp)	%-ID
(1) Contractile elements				
T080-F	Human myosin light chain 1, ventricular	emb|X07373|	247	97
T701-F	Human troponin T, cardiac	gb|L40162|	384	99
U117-F	Human actin, cardiac alpha	gb|J00068|	242	94
908-F	Human myosin light chain 2, ventricular	gb|S69022|	236	100
T834-F	Human myosin light chain 2, ventricular	gb|S69022|	290	95
T162-F	Human myosin light chain 2, cardiac	gb|M22815|	270	98
T061-F	Human myosin light chain 2, cardiac	gb|M22815|	380	99
T202-F	Rat calsequestrin	gb|U33287|	390	83
T656-F	Human troponin T, cardiac	emb|X74819|	350	98
T905-F	Human myosin heavy chain, cardic beta	gb|M21665|	243	97
T027-F	Human myosin heavy chain, cardic beta	gb|M25137|	170	78
T505-F	Human myosin heavy chain, cardic beta	emb|X05631|	240	98
T677-F	Human myosin heavy chain, cardic beta	gb|M21665|	188	96
T601-F	Human myosin alkali light chain, non-muscle	gb|M22918|	336	98
U198-F	Human myosin alkali light chain, ventricular	gb|M24248|	252	97
T808-F	Human myosin heavy chain, cardic beta	gb|M21665|	305	98
T203-F	Human myosin heavy chain, cardic beta	gb|M21665|	220	93
T773-F	Human myosin heavy chain, cardic beta	emb|X05631|	177	94
T076-F	Human myosin light chain 2, ventricular	gb|S69022|	260	92
T621-F	Human actin, alpha	emb|X00351|	197	100
U072-F	Human actin, cardiac alpha	emb|X03767|	153	95
T722-F	Human myosin regulatory light chain	emb|X54304|	196	94
T880-F	Human myosin regulatory light chain	emb|X54304|	417	98
T594-F	Human myosin regulatory light chain	emb|X54304|	301	86
T042-F	Human myosin light chain 2, ventricular	gb|S69022|	160	95
T133-F	Human troponin T, cardiac	gb|L40162|	263	100
T466-F	Human tropomyosin, skeletal alpha	gb|M19713|	395	96
T503-F	Human tropomyosin, skeletal alpha	gb|M19714|	269	98
U042-F	Human tropomyosin, skeletal alpha	gb|M19714|	165	93
T803-F	Human troponin C, slow skeletal	emb|X07897|	215	98
T726-F	Human troponin C, slow skeletal	emb|X07897|	178	98
T538-F	Human troponin C, slow skeletal	emb|X07897|	194	68
T721-F	Human troponin I, cardiac	emb|X54163|	236	97
T442-F	Human troponin T, cardiac	emb|X79857|	264	96
(2) Extracellular matrix				
T627-F	Human chondroitin/dermatan sulfate proteoglycan	gb|M14219|	202	99
T875-F	Human thrombospondin-4	emb|Z19585|	210	85
T861-F	Human gap junction protein, cardiac	emb|X52947|	157	100

(3) Cytoskeleton and related protein

U112-F	Human skeletal muscle 190 kDa protein	emb|X69090|	171	69
T095-F	Human desmin	gb|M63391|	390	100
U067-F	Human major nuclear matrix protein	gb|M63483|	230	97
T192-F	Human desmoplakin	gb|J05211|	165	96
T645-F	Human desmin	gb|M63391|	153	93
T205-F	Human desmin	gb|M63391|	280	99
T982-F	Human desmin	gb|M63391|	186	97
U171-F	Human crystallin, alpha B	gb|S45630|	347	96
T390-F	Human catenin, alpha	dbj|D13866|	179	97
U202-F	Human beta-sarcoglycan dystrophin-associated glycoprog	b|U29586|	192	97
T417-F	Human actinin, skeletal alpha-2	gb|M86406|	212	87
T003-F	Human titin	emb|X90568|	218	97
T995-F	Human titin	emb|X69490|	252	98
T472-F	Human titin	emb|X90568|	269	96
T469-F	Human titin	emb|X90568|	293	98
T450-F	Human titin	emb|X90568|	315	97
T449-F	Human titin	emb|X90568|	347	99
T711-F	Human actin depolymerization factor	gb|S65738|	385	98

(4) Energy metabolism

U043-F	Human cytochrome bc-1 complex core protein II	gb|J04973|	231	99
T665-F	Human pyruvate dehydrogenase E1-alpha subunit	gb|L48690|	325	98
U122-F	Human malate dehydrogenase, cytosolic	dbj|D55654|	167	95
T863-F	Human lipoprotein lipase	emb|X54516|	252	93
U176-F	Human glucose phosphate isomerase	gb|M55538|	167	98
T566-F	Human cytochrome c oxidase, subunit Va	gb|M22760|	175	85
U174-F	Human cytochrome c oxidase, subunit VIc	emb|X13238|	197	98
U134-F	Human cytochrome c oxidase, subunit VIIc	emb|X16560|	166	70
T345-F	Human cytochrome c oxidase, subunit VIIc	emb|X16560|	160	97
T760-F	Human cytochrome c oxidase, subunit VIIc	emb|X16560|	257	98
T980-F	Human cytochrome c oxidase, subunit VIIb	emb|Z14244|	176	97
U074-F	Human cytochrome c oxidase, subunit VIII	gb|J04823|	166	97
U203-F	Human cytochrome bc-1 complex core protein II	gb|J04973|	239	98
U048-F	Cow ATP synthetase, mitochondrial epsilon subunit	emb|M16978|	211	88
T218-F	Human creatine kinase, muscle	gb|M21494|	173	99
T256-F	Human creatine kinase, mitochondrial	gb|J05401|	235	100
T964-F	Human creatine kinase, mitochondrial	gb|J05401|	157	97
T299-F	Human calcium-ATPase	gb|M23115|	151	94
U159-F	Human ubiquinol cytochrome c reductase core I proteing	b|L16842|	277	98
T030-F	Human NADH-ubiquinone oxidoreductase	gb|M22538|	236	99
T187-F	Human F1-ATPase beta subunit	emb|X03559|	206	81
T865-F	Human ATPase coupling factor 6 subunit, mitochondrial	gb|M37104|	312	95
U057-F	Cow NADH ubiquinone oxidoreductase, MLRQ subunit	emb|X64897|	319	83
T684-F	Cow F1F0-ATP synthase complex F0 membrane domain g su	gb|S70448|	379	89

(5) Hormones and hormonal control

T596-F	Human tumor necrosis factor alpha inducible protein A	gb|M59465|	355	98
T424-F	Human brain natriuretic protein	gb|M25296|	279	99
T715-F	Human cardiodilation atrial natriuretic factor	gb|M30262|	295	93
T094-F	Human cardiodilation atrial natriuretic factor	gb|M30262|	334	97
T929-F	Human cardiodilation atrial natriuretic factor	gb|M30262|	206	98
T064-F	Human cardiodilation atrial natriuretic factor	gb|I01397|	192	99
T811-F	Human interferon-inducible gene 1-8U	emb|X57352|	250	99
T795-F	Human prolactin receptor-associated protein	gb|M18981|	313	93
T091-F	Human prostaglandin D synthase	gb|M61900|	215	92
T777-F	Rat vascular endothelial growth factor	gb|U22372|	303	98

(6) Membrane associated

T262-F	Rat integral membrane glycoprotein	emb|Z21513|	310	87
T812-F	Human leukemia virus receptor 1	gb|L20859|	256	97
T706-F	Human glycosylated surface protein	emb|X06296|	338	99
U128-F	Human p190-B	gb|U17032|	247	92
U212-F	Human nucleoporin-like protein	emb|X89478|	220	98
T928-F	Human nucleoporin NUP358	gb|L41840|	383	100
T989-F	Human nucleoporin NUP358	dbj|D42063|	186	98
T352-F	Human glutamate receptor flip isoform	gb|U10301|	163	94
T569-F	Human axonal transporter	emb|X90840|	200	76
T910-F	Human myoblast cell surface antigen 24	emb|X16850|	222	91
T628-F	Human membrane glycoprotein 4F2 antigen heavy chain	gb|J02939|	345	97

(7) Miscellaneous

T673-F	Human merosin	gb|M59832|	337	98
T870-F	Human homolog of rat HREV107-like protein	emb|X92814|	308	98
T766-F	Human histidine-rich calcium binding protein	gb|M60052|	401	97
T573-F	Human moesin	gb|M69066|	180	98
T841-F	Human neurodap1	dbj|D32249|	357	91
T878-F	Human platelet-endothelial tetraspan antigen 3	gb|U14650|	160	98
T710-F	Human porin	gb|L08666|	316	97
T901-F	Rat nuclear pore complex protein NUP107	gb|L31840|	380	85
T576-F	Human supressor for yeast mutant	dbj|D66904|	201	89
U216-F	Human highly basic protein, 23 kDa	emb|X56932|	231	96
T509-F	Human growth factor inducible intermediate	emb|X56790|	321	75
T888-F	Human visinin-like peptide 1 homolog	gb|U14747|	260	84
T660-F	Human vinculin	gb|M33308|	323	97
T679-F	Human decorin	gb|M98262|	171	95
T616-F	Human calcyclin	gb|J02763|	379	99
T132-F	Human Tob	dbj|D38305|	265	100
U051-F	Human OXA1Hs	emb|X80695|	208	94
T216-F	Human Ki nuclear autoantigen	gb|U11292|	365	98
T272-F	Human ETS2	emb|X55181|	208	95
T741-F	Human DS-1	emb|X81788|	317	98
T599-F	Human DGCR6	emb|X96484|	293	95
T001-F	Human Csa-19	gb|U12404|	266	100
T904-F	Human CAC and GTG repeat-containing mRNA	gb|U00943|	246	97
T160-F	Human BTG1	emb|X61123|	316	99

(8) Other metabolism

T313-F	Human metallothionein I-G	gb|J03910|	295	97
U148-F	Human inositol phospholipid assembly protein	gb|S74936|	310	98
T658-F	Human long-chain acyl-Co-A synthetase	dbj|D10040|	331	95
U221-F	Human lysyl synthetase	dbj|D31890|	205	99
T997-F	Human macroglobulin, beta-2	emb|X07621|	168	91
U071-F	Mous inosine-5'-monophosphate dehydrogenase	gb|M98333|	213	98
T226-F	Human manganese superoxide dismutase	emb|X65965|	182	86
T147-F	Human metallothionein I-A	gb|K01383|	250	89
T058-F	Human glycogen 4-alpha-D-glycosytransferase	gb|U32573|	415	91
T923-F	Human metallothionein I-e	gb|M10942|	190	100
T327-F	Human metallothionein II	gb|M26637|	170	95
T626-F	Human metallothionein II	gb|M26637|	305	92
T994-F	Human metallothionein from cadium-treated cells	emb|V00594|	197	100
T818-F	Human metallothionein from cadium-treated cells	emb|V00594|	220	94
T954-F	Human nicotinamide nucleotide transhydrogenase	gb|U40490|	154	95

U135-F	Human proteasome subunit HsN3	dbj|D26600|	222	98
U115-F	Human short chain acyl-Co-A dehydrogenase	gb|M26393|	154	89
T735-F	Human 26S protease regulatory subunit	gb|L02426|	358	98
T877-F	Human ubiquitin	gb|M26880|	223	100
T025-F	Human ATP-citrate lyase	emb|X64330|	174	100
U214-F	Human tissue inhibitor of metalloproteinases-3	gb|U14394|	164	93
T211-F	Human alpha-2-macroglobulin receptor-associated prote	gb|M63959	392	96
T836-F	Human ubiquitin	gb|M26880|	238	97
T814-F	Human branched chain acyltransferase	gb|J03208|	213	98
T006-F	Human calpastatin	gb|U31345|	253	98
U164-F	Human carbonic anhydrase III	gb|M29453|	204	96
T680-F	Human ferritin light chain	gb|M11147|	188	98
T686-F	Human cathepsin B	gb|L22569|	360	97
T201-F	Human cathepsin L	gb|L06426|	190	82
T736-F	Human ubiquitin	gb|M26880|	352	98
T553-F	Human ubiquitin-like protein	dbj|D23662|	215	98
T708-F	Human very-long-chain acyl-CoA dehydrogenase	dbj|D43682|	303	100
T102-F	Human ferritin heavy chain	gb|M12937|	282	98
U215-F	Human ferritin heavy chain	gb|M12937|	300	99
T692-F	Human ferritin heavy chain	gb|M15383|	156	98
U183-F	Human myoglobin	emb|X00373|	219	99
T143-F	Human plasma protein S, vitamin K-dependent	gb|M15036|	413	99

(9) Signal transduction / cell regulation

T909-F	Human N-ras	emb|X02751|	240	93
T141-F	Human protein kinase C inhibitor-I	gb|U27143|	452	98
T204-F	Human pyruvate kinase, M-2 type	gb|M23725|	234	93
T036-F	Rat Enigma	gb|U48247|	210	90
T854-F	Rat protein tyrosin kinase JAK2	gb|U13396|	152	89
T115-F	Human serine/threonine kinase receptor	gb|U20165|	294	100
T052-F	Human serine/threonine kinase receptor	gb|U20165|	193	99

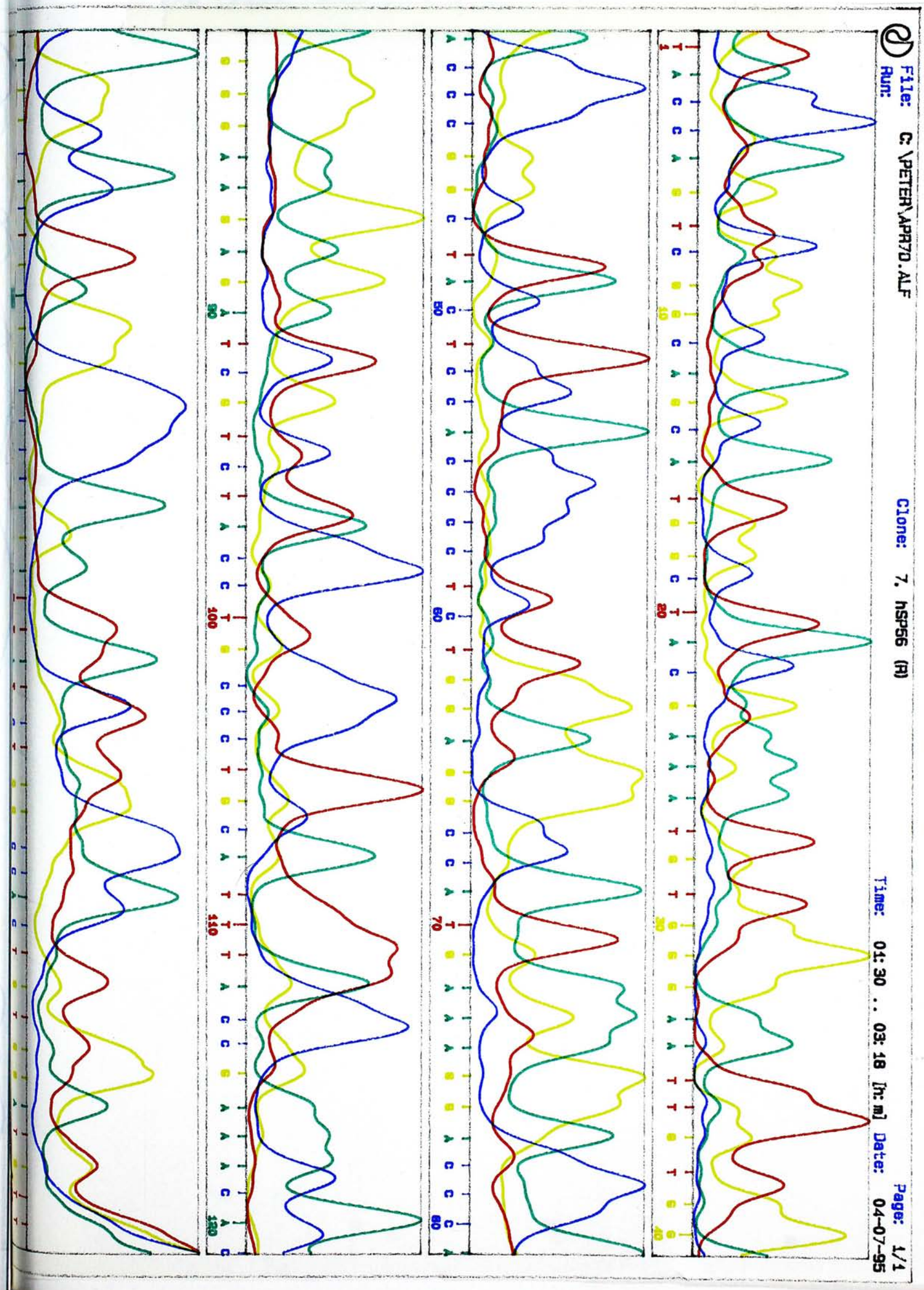
(10) Transcription and translation

T620-F	Human ribosomal protein S10	gb|U14972|	570	98
U100-F	Human ribosomal protein L27a	gb|U14968|	238	100
T552-F	Human ribosomal protein L27a	gb|U14968|	163	91
T078-F	Human ribosomal protein L30	gb|M94314|	236	98
U110-F	Human ribosomal protein L30	emb|X79238|	247	93
T548-F	Human ribosomal protein L37a	gb|L22154|	235	99
U194-F	Human ribosomal protein L39	emb|X82551|	194	88
T200-F	Human ribosomal protein L4	dbj|D23660	242	81
T380-F	Human ribosomal protein L26	emb|X69392|	160	95
T916-F	Human ribosomal protein S14	gb|M13934|	290	99
T784-F	Human ribosomal protein S16	gb|M60854|	330	99
U211-F	Human ribosomal protein S18	emb|X69150|	177	98
T667-F	Human ribosomal protein S21	gb|L04483|	222	94
T748-F	Human ribosomal protein S4X	gb|M58458|	210	99
T937-F	Human ribosomal protein S4X	gb|M58458|	188	98
T944-F	Human ribosomal protein S4X	gb|M58458|	263	100
T461-F	Human ribosomal protein S5	gb|U14970|	239	96
U205-F	Human ribosomal prtoein L9	gb|U21138|	290	97
T255-F	Human U1 small nuclear RNP-specific C protein	emb|X12517|	252	96
T153-F	Human heat shock protein, 90 kDa	gb|M16660|	300	86
T009-F	Human acidic ribosomal phosphoprotein P0	gb|M17885|	250	97
T753-F	Human acidic ribosomal phosphoprotein P2	gb|M17887|	215	98
T648-F	Human zinc finger protein, bcl-6	gb|U00115|	205	98
T051-F	Human elongation factor 1-alpha	gb|L10340|	194	97
U061-F	Human elongation factor 1-alpha	gb|M29548|	233	95
U195-F	Human heat shock protein, 70 kDa	emb|X04676|	270	98
T264-F	Human heat shock protein, 70B'	emb|X51758|	266	99
T065-F	Human heat shock protein, 71 kDa	emb|Y00371|	304	96
T321-F	Human ribosomal protein L26	emb|X69392|	294	96
U032-F	Human hnRNP C2	gb|M29063|	284	96
T623-F	Human hnRNP E1	gb|X78137|	249	97
T681-F	Human homolog of Drosophila enhancer of split	gb|U04241|	174	98
T720-F	Mouse 60S ribosomal protein	gb|U28917|	218	85
T771-F	Human nuclear factor p97	gb|L39793|	341	88
T151-F	Human ras-like protein	gb|M31468|	278	92
T675-F	Human ribosomal protein L18	gb|L11566|	289	98
U153-F	Human ribosomal protein L23a	gb|U02032|	327	96

3.3 Sequence analysis of hSP56

Among the 2244 human fetal heart cDNA EST clones that were sequenced by myself and my colleagues (Hwang *et al.,* 1995), one of the human homologues has its DNA sequence resembling mSP56. It was named human 56 kDa selenium-binding protein (hSP56). The complete cDNA sequence was obtained by primer walking (primers sequences 8a-8h are listed in Appendix 1). The sequence printout of the cDNA from the automated sequencing machine named Pharmacia ALF sequencer is shown in Fig 3-1. The various peaks represent the various bases, which are shown in different colours. The DNA sequence of the cDNA clones determined both by cycle sequencing and manual sequencing (Tsui *et al.*, 1994) was shown in Fig.3-2. The size of the insert as determined by gel electrophoresis of PCR product was about 1440 bp and is shown in Fig 3-3. Excluding the vector sequences at both ends of the cDNA clone and the poly A region, the calculated length of the cDNA insert was 1419 bp. There is only one hSP56 cDNA sequence amongst

Fig. 3-1. A printout of the DNA sequence of hSP56 cDNA obtained from ALF Manager. It shows the first 160 base pairs of the full length hSP56 sequence.



the 2244 fetal heart cDNA clones sequenced in our laboratory. When sent to GenBank for alignment, the cDNA clone was found to be very similar to a mouse gene which is named mouse selenium-binding protein. It is a cDNA which was first found in the liver of mouse and the DNA sequence of the cDNA clone was introduced in Chapter 1.4. The cDNA clone which we obtained was named human 56 kDa selenium-binding protein (hSP56) as we first identified it in a human fetal heart and it was highly homologues with mouse 56 kDa selenium-binding protein (mSP56). To study the possible open reading frame (ORF) of hSP56, an analysis using DNASIS™ was performed (Fig. 3-2). There are a number of possible ORFs for the cDNA sequence. It was found that one of those amino acid sequence encoded is very similar to that of mouse hSP56. The alignment of mSP56 with hSP56 suggested that the probable initiation codon is located at nucleotide number 15 and the stop codon is located at nucleotide number 1434. The 5' UTR has a length of 16 basepairs and the 3'UTR has a length of 234 base pairs. Although no 5' primer extension experiment is done, the 5'

Fig. 3-2. The DNA and amino acid sequence of hSP56. The start codon (ATG), stop codon (TGA), and polyadenylation signal (AATAAA) are underlined.

```
1                                M   A   T   K   C   G   N   C   G   P   G      11
1   TGT ACC AGT CGC AGC ATG GCT ACG AAA TGT GGG AAT TGT GGA CCC GGC             48

12  Y   S   T   P   L   E   A   M   K   G   P   R   E   E   I   V      27
49  TAC TCC ACC CCT CTG GAG GCC ATG AAA GGA CCC AGG GAA GAG ATC GTC             96

28  Y   L   P   C   I   Y   R   N   T   G   T   E   A   P   D   Y      43
97  TAC CTG CCC TGC ATT TAC CGA AAC ACA GGC ACT GAG GCC CCA GAT TAT            144

44  L   A   T   V   D   V   D   P   K   S   P   Q   Y   C   Q   V      59
145 CTG GCC ACT GTG GAT GTT GAC CCC AAG TCT CCC CAG TAT TGC CAG GTC            192

60  I   H   R   L   P   M   P   N   L   K   D   E   L   H   H   S      75
193 ATC CAC CGG CTG CCC ATG CCC AAC CTG AAG GAC GAG CTG CAT CAC TCA            240

76  G   W   N   T   C   S   S   C   F   G   D   S   T   K   S   R      91
241 GGA TGG AAC ACC TGC AGC AGC TGC TTC GGT GAT AGC ACC AAG TCG CGC            288

92  N   K   L   V   L   P   S   L   I   S   S   R   I   Y   V   V     107
289 AAC AAG CTG GTC CTG CCC AGT CTC ATC TCC TCT CGC ATC TAT GTG GTG            336

108 D   V   G   S   E   P   G   P   Q   K   L   H   K   V   I   E     123
337 GAC GTG GGC TCT GAG CCC GGG CCC CAA AAG CTG CAC AAG GTC ATT GAG            384

124 P   K   D   I   H   A   K   C   E   L   A   C   L   H   T   S     139
385 CCC AAG GAC ATC CAT GCC AAG TGC GAA CTG GCC TGT CTC CAC ACC AGC            432

140 H   C   L   A   S   G   E   V   M   I   S   S   L   G   D   V     155
433 CAC TGC CTG GCC AGC GGG GAA GTG ATG ATC AGC TCC CTG GGG GAC GTC            480

156 K   G   N   G   K   G   G   F   V   L   L   D   G   E   T   F     171
481 AAG GGC AAT GGC AAA GGG GGT TTT GTG CTG CTG GAT GGG GAG ACG TTC            528

172 E   V   K   G   T   W   E   R   P   G   G   A   A   P   L   G     187
529 GAG GTG AAG GGG ACA TGG GAG AGA CCT GGG GGT GCT GCA CCG TTG GGC            576

188 Y   D   F   W   Y   Q   P   R   H   N   V   M   I   S   T   E     203
577 TAT GAC TTC TGG TAC CAG CCT CGA CAC AAT GTC ATG ATC AGC ACT GAG            624

204 W   A   A   P   N   V   L   R   D   G   F   N   P   A   D   V     219
625 TGG GCA GCT CCC AAT GTC TTA CGA GAT GGC TTT AAC CCC GCT GAT GTG            672

220 E   A   G   L   Y   G   S   H   L   Y   V   W   D   W   Q   R     235
673 GAG GCT GGA CTG TAC GGG AGC CAC TTA TAT GTA TGG GAC TGG CAG CGC            720

236 H   E   I   V   Q   T   L   S   L   K   D   G   L   I   P   L     251
721 CAT GAG ATT GTG CAG ACC CTG TCT CTA AAA GAT GGG CTG ATA CCC TTG            768

252 E   I   R   F   L   H   N   P   S   A   T   Q   G   F   V   G     267
769 GAG ATC CGC TTC CTG CAC AAC CCA AGT GCC ACC CAG GGT TTT GTA GGC            816

268 C   A   S   A   P   N   I   Q   R   F   Y   K   T   R   E   G     283
817 TGT GCC TCA GCT CCA AAC ATC CAG CGC TTC TAC AAA ACG AGG GAA GGT            864

284 T   W   S   V   E   K   V   I   Q   V   P   P   K   K   V   K     299
865 ACA TGG TCA GTG GAG AAG GTG ATC CAG GTG CCC CCC AAG AAA GTG AAG            912

300 G   W   L   L   P   G   V   P   G   L   I   T   D   I   L   L     315
913 GGC TGG CTG CTG CCA GGG GTG CCA GGC CTG ATC ACC GAC ATC CTG CTC            960
```

con't Fig. 3-2.

```
316   S   L   D   D   R   F   L   Y   F   S   N   W   L   H   G   D    331
961   TCC CTG GAC GAC CGC TTC CTC TAC TTC AGC AAC TGG CTG CAT GGG GAC   1008

332   L   R   Q   Y   D   I   S   D   P   Q   R   P   R   L   T   G    347
1009  CTG AGG CAG TAT GAC ATC TCT GAC CCA CAG AGA CCC CGC CTC ACA GGA   1056

348   Q   L   F   L   G   G   S   I   V   K   G   G   P   V   Q   V    363
1057  CAG CTC TTC CTC GGA GGC AGC ATT GTT AAG GGA GGC CCT GTG CAA GTG   1104

364   L   E   D   E   E   L   K   S   Q   P   E   P   L   V   V   K    379
1105  CTG GAG GAC GAG GAA CTA AAG TCC CAG CCA GAG CCC CTA GTG GTC AAG   1152

380   G   K   R   V   A   G   G   P   Q   M   I   Q   L   S   L   D    395
1153  GGA AAA CGG GTG GCT GGA GGC CCT CAG ATG ATC CAG CTC AGC CTG GAT   1200

396   G   K   R   L   Y   I   T   T   S   L   Y   S   A   W   D   K    411
1201  GGC AAG CGC CTC TAC ATC ACC ACG TCG CTG TAC AGT GCC TGG GAC AAG   1248

412   Q   F   Y   P   D   L   I   R   E   G   S   V   M   L   Q   V    427
1249  CAG TTT TAC CCT GAT CTC ATC AGG GAA GGC TCT GTG ATG CTG CAG GTT   1296

428   D   V   D   T   V   K   G   G   L   K   L   N   P   N   C   L    443
1297  GAT GTA GAC ACA GTA AAA GGA GGG CTG AAG TTG AAC CCC AAC TGC CTG   1344

444   V   D   F   G   K   E   P   L   G   P   A   L   A   H   E   L    459
1345  GTG GAC TTC GGG AAG GAG CCC CTT GGC CCA GCC CTG GCT CAC GAG CTT   1392

460   R   Y   P   G   G   D   C   S   S   D   I   W   I   *            473
1393  CGC TAC CCT GGG GGC GAT TGT AGC TCT GAC ATC TGG ATT TGA ACT CCA   1440

1441  CCC TCA TCA CCC ACA CTC CCT ATT TTG GGC CCT CAC TTC CTT GGG GAC   1488

1489  CTG GCT TCA TTC TGC TCT CTC TTG GCA CCC GAC CCT TGG CAG CAT GTA   1536

1537  CCA CAC AGC CAA GCT GAG ACT GTG GCA ATG TGT TGA GTC ATA TAC ATT   1584

1585  TAC TGA CCA CTG TTG CTT GTT GCT CAC TGT GCT GCT TTT CCA TGA GCT   1632

1633  CTT GGA GGC ACC AAG AAA TAA ACT CGT AAC CCT GTC                  1668
```

Fig. 3-3. The PCR product of hSP56. 'M' represents λ/*Hind*III-ϕX174/*Hae*III size marker. The 'PCR' indicates the target PCR product of hSP56 using pAED4-hSP56 as the template. The primers are sequences as described in page 50 of the materials and methods section.

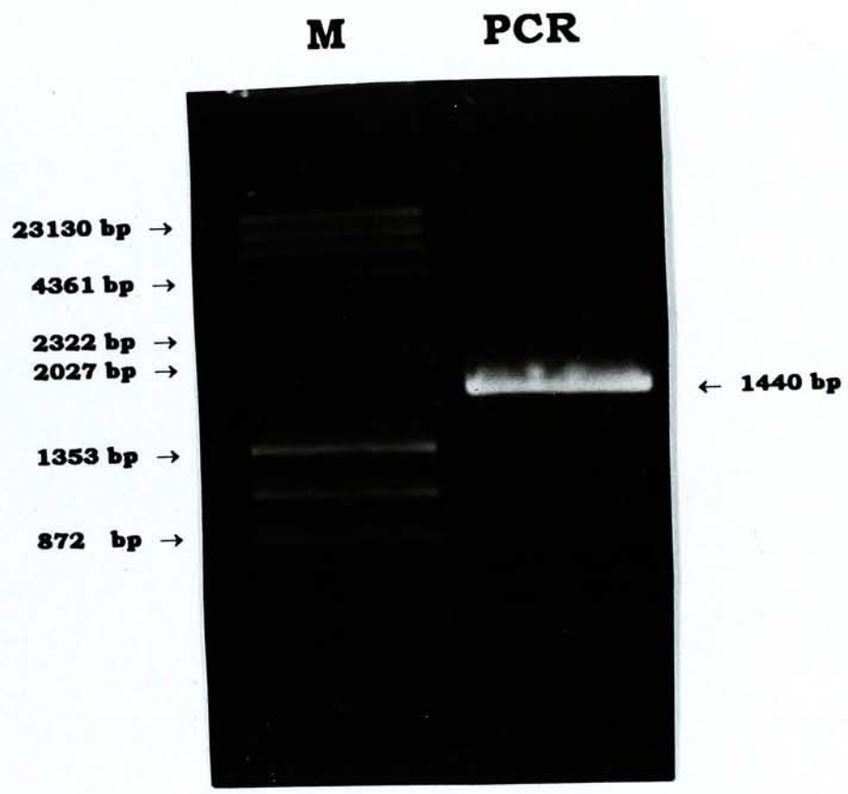
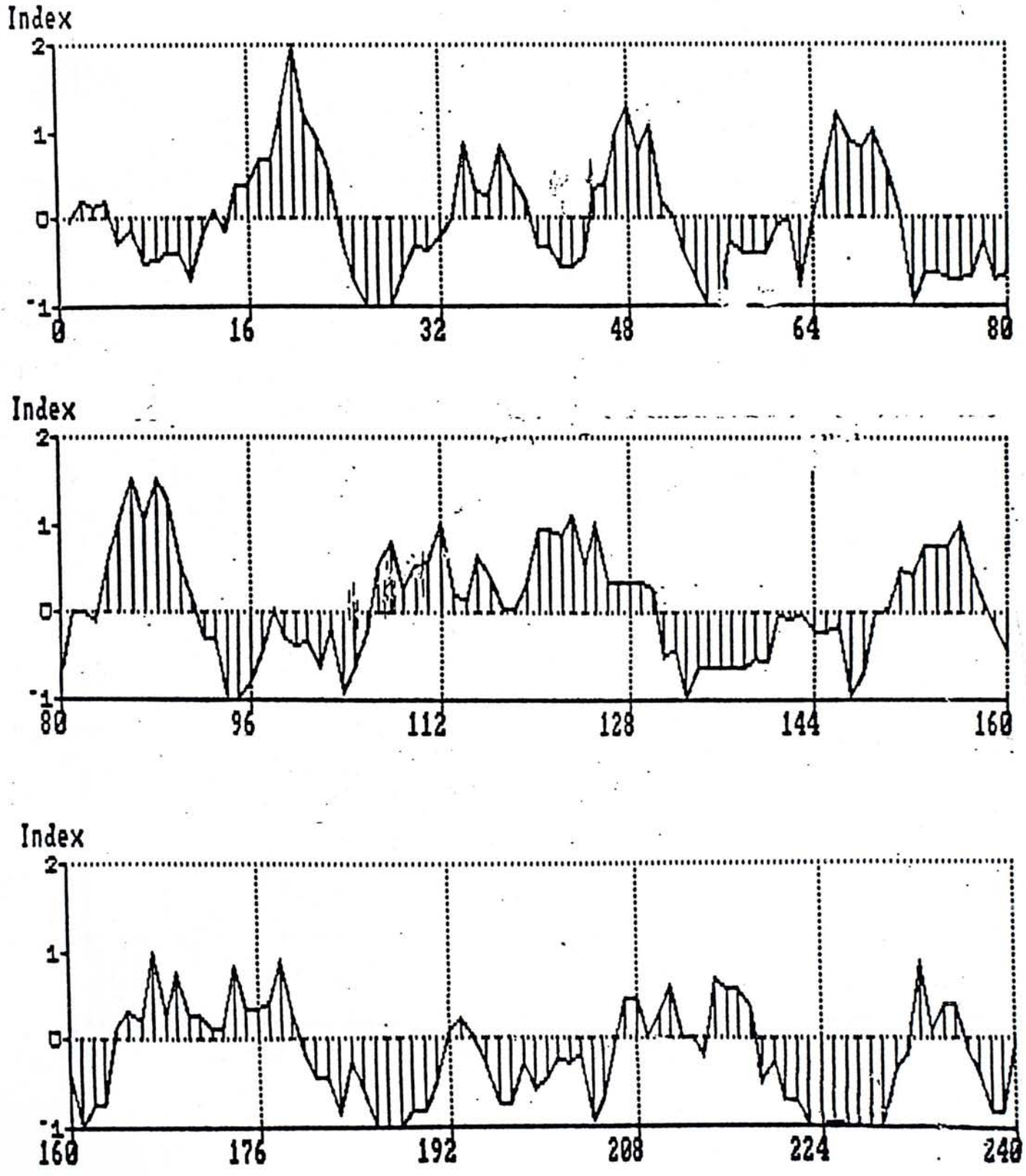
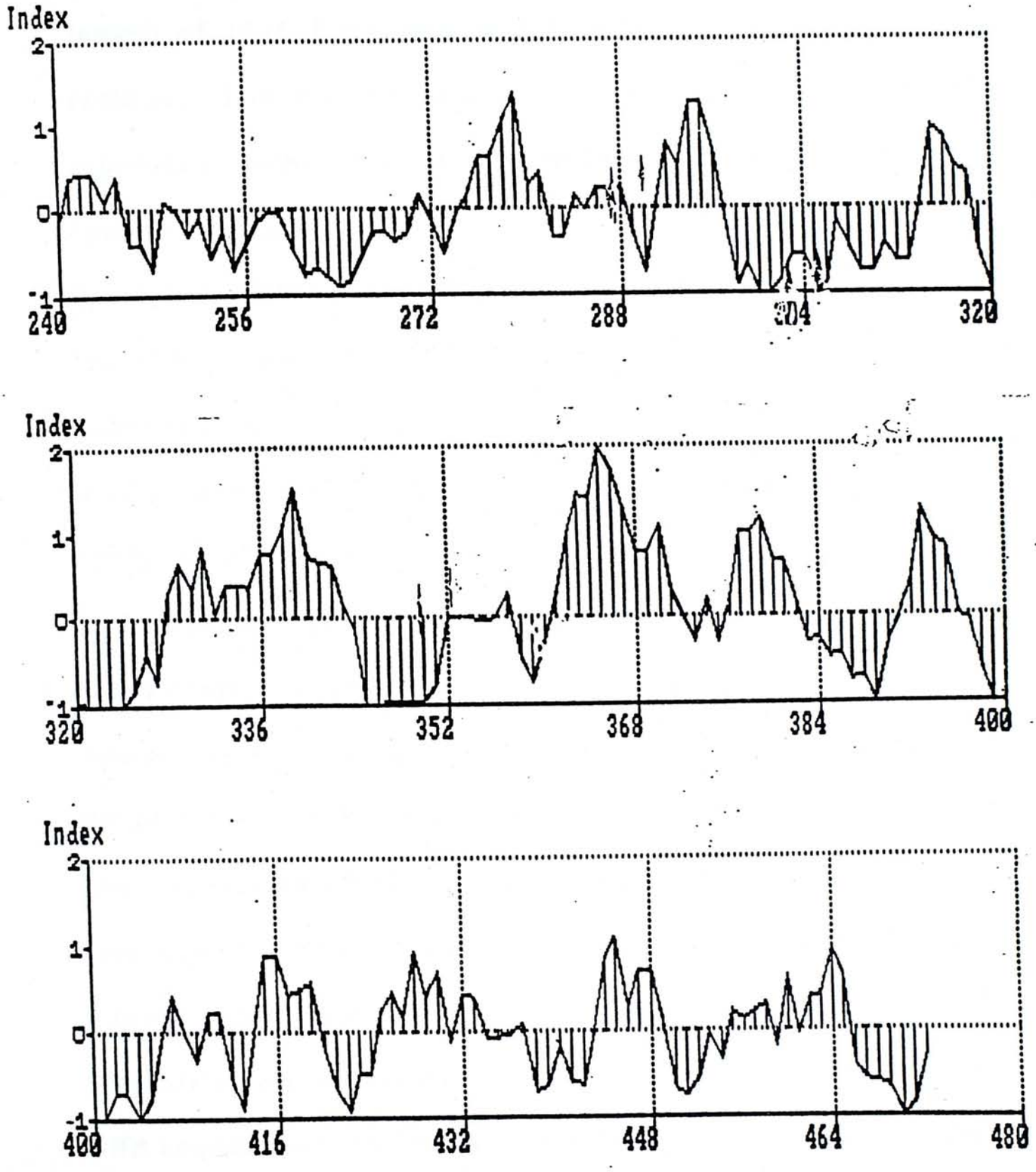


Fig. 3-4. Hydrophobicity analysis of hSP56 as determined by PROSIS™: (Hopp.Thr). The averaging "window" size for analysis was set as six.



con't. Fig. 3-4.


Index
2
1
0
-1
240
256
272
288
304
320
Index
2
1
0
-1
320
336
352
368
384
400
Index
2
1
0
-1
400
416
432
448
464
480

non coding sequence is 99% similar to the mSP56, therefore hSP56 is likely to be complete. The ORF of hSP56 has a length of 1419 base pairs which codes for 472 amino acid residues. The consensus initiation sequence RCCAUGG (R represent purine) is present at the start of the ORF and the typical polyadenylation signal, AATAAA (Proudfoot and Brownlee, 1977) is found at the 3' UTR of the clone. After translating the ORF of hSP56, a gene product of 472 amino acid residues was obtained with the estimated size of 52250 dalton as estimated by DNASIS™ . From the hydrophobicity analysis of PROSIS™, the protein has no prominent hydrophobic region (Fig.3-4) and so it was unlikely to be a transmembrane protein. The amino acid content of hSP56 was shown in Table 3-2. The estimated pI as determined by PROSIS™ was 6.13 (Fig. 3-5). The DNA sequence of hSP56 has a similarity of 87.2 % when aligned with that of mSP56, respectively. When the amino acid sequence of hSP56 was aligned with that of mSP56, 87.3 % identity, 98.7 % similarity, respectively, were found. The alignment of the DNA sequences between the two cDNA clones is shown in Fig.

3-6. From the comparison of DNA and amino acid sequences of hSP56 and mSP56, It is highly probable that hSP56 and mSP56 are structurally similar to each other because they have a high degree of amino acid sequence identity. We proposed that hSP56 and mSP56 are homologs in different organisms.

Table.3-4. Table of amino acid content of hSP56 as analysed by DNASIS™.

Amino Acid		Mol%	Count (CNT)	Residue Weight (RW)	CNT * RW
Gly	G	9.96	47	57.05	2681.35
Ala	A	4.24	20	71.07	1421.40
Val	V	7.42	35	99.13	3469.55
Leu	L	11.02	52	113.15	5883.80
Ile	I	4.87	23	113.15	2602.45
Ser	S	6.78	32	87.07	2786.24
Thr	T	4.24	20	101.10	2022.00
Cys	C	2.54	12	103.14	1237.68
Met	M	1.48	7	131.19	918.33
Asp	D	5.93	28	115.08	3222.24
Asn	N	2.97	14	114.10	1597.40
Glu	E	5.51	26	129.11	3356.86
Gln	Q	3.81	18	128.13	2306.34
Arg	R	4.24	20	156.18	3123.60
Lys	K	5.93	28	128.17	3588.76
His	H	2.75	13	137.14	1782.82
Phe	F	2.75	13	147.17	1913.21
Tyr	Y	3.60	17	163.17	2773.89
Trp	W	2.33	11	186.20	2048.20
Pro	P	7.63	36	97.11	3495.96

TOTAL RESIDUE WEIGHT -> 52232.08

+) 18.02

MOLECULAR WEIGHT -----> 52250.10

Fig. 3-5. Estimated pI of hSP56 as determined by PROSIS™.

```
THE NUMBER OF CHARGED AMINO ACID
  Arg (R) :      20    pKa = 12.5
  Asp (D) :      28    pKa =  3.9
  Cys (C) :      12    pKa =  8.3
  Glu (E) :      26    pKa =  4.3
  His (H) :      13    pKa =  6.0
  Lys (K) :      28    pKa = 10.5
  Tyr (Y) :      17    pKa = 10.1

  N_terminal Amino Acid : Met (M)    pKa =  9.3
  C_terminal Amino Acid : Ile (I)    pKa =  2.3

  Isoelectric point [pI] =  6.13
```

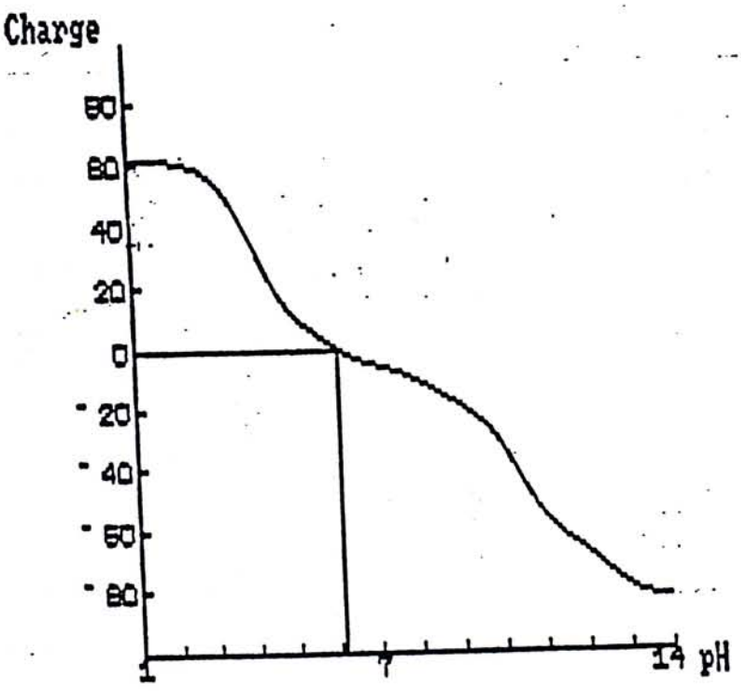


Fig. 3-6. Comparison of DNA sequence of hSP56 and mSP56. The alignment was performed using MacDNASIS™. Amino acids that are identical between the two sequences are marked by ':',that are similar are marked by '.'. while those that are different are left blank.

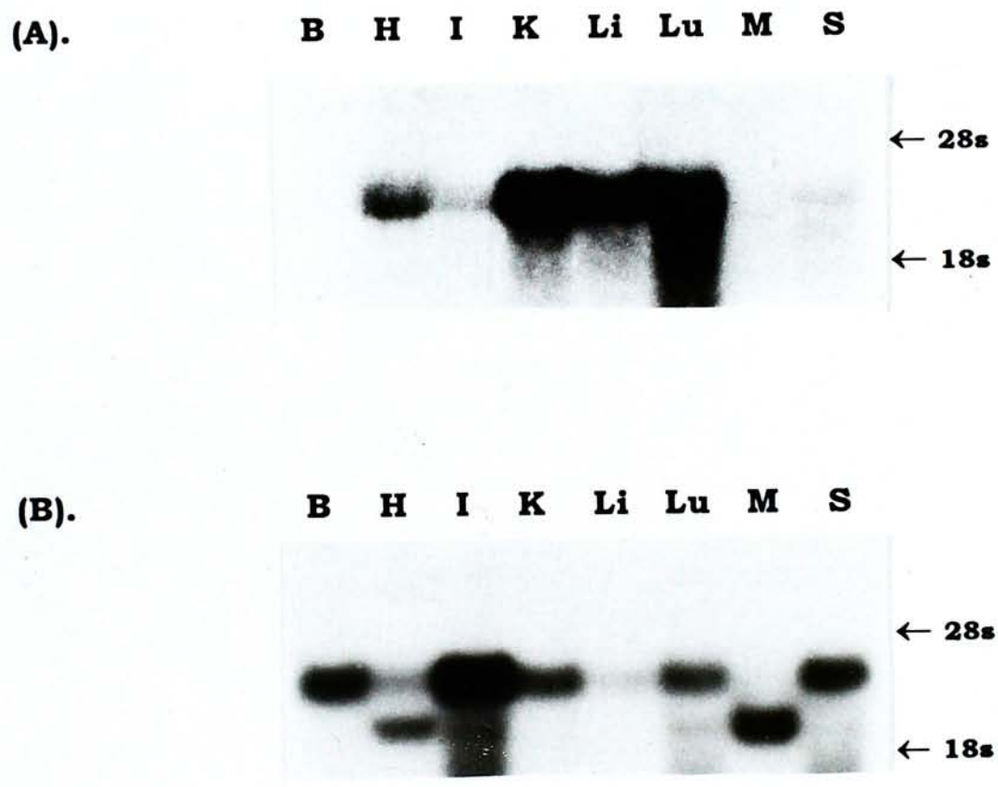
```
                10        20        30        40        50        60
hSP56   MATKCGNCGPGYSTPLEAMKGPREEIVYLPCIYRNTGTEAPDYLATVDVDPKSPQYCQVI
        ::::::.:::::::::::::::::::::::::::::::::::::::::::::::::.:::
mSP56   MATKCTKCGPGYSTPLEAMKGPREEIVYLPCIYRNTGTEAPDYLATVDVDPKSPQYSQVI

                70        80        90        100       110       120
hSP56   HRLPMPNLKDELHHSGWNTCSSCFGDSTKSRNKLVLPSLISSRIYVVDVGSEPGPQKLHK
        :::::: :::::::::::::::::::::::::::.::.:::::::::::::::: ..:::::
mSP56   HRLPMPYLKDELHHSGWNTCSSCFGDSTKSRNKLILPGLISSRIYVVDVGSEPRAPKLHK

                130       140       150       160       170       180
hSP56   VIEPKDIHAKCELACLHTSHCLASGEVMISSLGDVKGNGKGGFVLLDGETFEVKGTWERP
        :::...:.:::....:::::::::::::.:.::::..:::::.::::::::::::::::.:
mSP56   VIEASEIQAKCNVSSLHTSHCLASGEVMVSTLGDLQGNGKGSFVLLDGETFEVKGTWEKP

                190       200       210       220       230       240
hSP56   GGAAPLGYDFWYQPRHNVMISTEWAAPNVLRDGFNPADVEAGLYGSHLYVWDWQRHEIVQ
        :.::::::::::::::::.:::::::::..:::::.:::::::::...:::::::::.:
mSP56   GDAAPMGYDFWYQPRHNVMVSTEWAAPNVFKDGFNPAHVEAGLYGSRIFVWDWQRHEIIQ

                250       260       270       280       290       300
hSP56   TLSLKDGLIPLEIRFLHNPSATQGFVGCASAPNIQRFYKTREGTWSVEKVIQVPPKKVKG
        :: ..:::::::::::::.:::::::::::::::::::::::. ::::::::::::::::::
mSP56   TLQMTDGLIPLEIRFLHDPSATQGFVGCASAPNIQRFYKNAEGTWSVEKVIQVPSKKVKG

                310       320       330       340       350       360
hSP56   WLLPGVPGLITDILLSLDDRFLYFSNWLHGDLRQYDISDPQRPRLTGQLFLGGSIVKGGP
        :.:::::::::::::::::::::::::::::::.:::::::.::.:::.::.::::::::.::.
mSP56   WMLPGVPGLITDILLSLDDRFLYFSNWLHGDIRQYDISNPQKPRLAGQIFLGGSIVRGGS

                370       380       390       400       410       420
hSP56   VQVLEDEELKSQPEPLVVKGKRVAGGPQMIQLSLDGKRLYITTSLYSAWEKQFYPDLIRE
        ::::::.::..:::::::::::::..:::::::::::::::: :::::::::.::::::::::
mSP56   VQVLEDQELTCQPEPLVVKGKRIPGGPQMIQLSLDGKRLYATTSLYSAWDKQFYPDLIRE

                430       440       450       460       470
hSP56   GSVMLQVDVDTVKGGLKLNPNCLVDFGKEPLGPALAHELRYPGGDCSSDIWI
        ::.::::.::::::.:::::::: :::::::::::.::::::::::::::::::::
mSP56   GSMMLQIDVDTVNGGLKLNPNFLVDFGKEPLGAALAHELRYPGGDCSSDIWI
```

3.4 Northern hybridization of hSP56

Since the homology between hSP56, mSP56 and is very high, the hSP56 probe was used to hybridize with total RNA of various mouse tissues (Fig.3-7). Total RNA was isolated from mouse organs using acid guanidium thiocyanate phenol-chloroform extraction. By comparing the location of the band with that of 18S and 28S ribosomal RNAs, it was found that the hSP56 mRNAs had the size of about 1660 nucleotides. Since mature eukaryotic mRNAs have a poly A tail of 200-300 adenine residues at its 3' end (Wickens, 1990), the size of the cDNA in the hSP56 clone (about 1440 nucleotides) would be close to the size of the mature mSP56 mRNA. For the normalization of the hSP56 signal, β-actin cDNA was used as a control probe. To confirm whether we have obtained a full length clone which possessed a full length 5' UTR, primer extension experiments was performed to answer the question. It was shown that liver, lung, and kidney have the highest signal. A moderate signal could be seen in heart ; detectable, but low, signals were found in

intestine and spleen. Virtually no signal could be detected in skeletal muscle and brain. Previous results have shown that SP56 is expressed at a high level in mouse liver, kidney, and lung, with a lower level in intestine and barely detectable amounts in brain, thymus, muscle, spleen, skin, mammary tissues, testis, and ovary (Lanfear *et al.*, 1993; Morrison *et al.*, 1989). Such previous results agree well with ours. However, here we further show that the expression of SP56 in mouse heart is at an intermediate level between those found in liver/lung/kidney and intestine.

Fig. 3-7. Northern hybridization. (A): Key for mouse RNA sample: B, brian; H, heart; I, intestine; K, kidney; Li, liver; Lu, lung; M, muscle; S, spleen. (B): β-actin for normalization



3.5 Cloning of hSP56 into pAED4

Before the expression of hSP56 in *E coli*, we need first ensure that the insert cloned into the vector is at an optimal distance to a promoter upstream and is in a correct reading frame with respect to the start codon of the vector. Therefore, in order to produce the authentic hSP56, we have designed a primer which was complementary to the 5' end of the ORF. This primer was designed such that the ATG of the ORF of hSP56 was able to ligate with the *NdeI* site of pAED4. When hSP56 was ligated with pAED4 through the *NdeI* restriction site, hSP56 was under the control of the T7 promoter of the vector. At the 3' end, the oligo dT primer would prime along the poly A tail of the cDNA clone. The cDNA insert that is amplified by PCR is then restricted by *NdeI* and *Eco*RI and then ligated to the vector pAED4 for expression. By using this method, hSP56 was successfully amplified. It can be shown that the PCR product of hSP56 using cloning and dT primer was of a smaller size when compared with that using the common PCR primers flanking

the cloning site of λgt22, because the cloning and oligo dT primers would remove the vector sequences at 5' and 3' end, for example, the 5'UTR and a part of the poly A tail of the original hSP56 cDNA clone. The expected size of the PCR product was about 1420 bp. There was a range of PCR products due to the fact that the dT primer could prime along different sites of the poly A region of the 3' end of the clone. Individual cDNA clone can be isolated after transformation into *E. coli.*

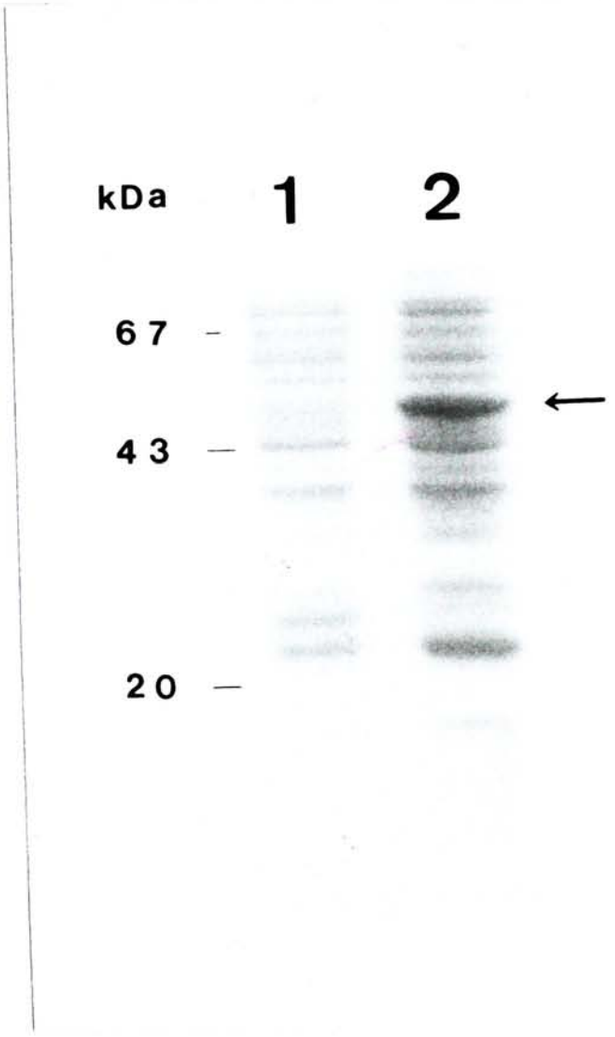
3.6 Expression of the hSP56 in E.coli

The oligo-dT primer can be used as a common 3' end cloning primer to amplify the interesting cDNA clones found in our human heart cDNA sequencing project. In my study project, hSp56 cDNA was successfully amplified using a tailor-made cloning primer and an oligo dT primer. This strategy is suitable for large scale cDNA sequencing and expression project. Some of my coworkers have used this strategy to clone and express the genes successfully. The former cDNA was amplified from the adult and fetal heart cDNA libraries while the latter cDNA was amplified from two overlapping clones in the adult heart cDNA library. Their successes proved that this strategy is very useful and efficient. The success of the directional cloning was proved by restriction cutting of the putative recombinant plasmid.

After transformation, some of the clones which were able to grow in LBAC plates were checked for there ability to express hSP56. BL21 (DE3) pLysS and BL21 (DE3) pLysE

were used as the host. In my experiment, all 12 clones of BL21(DE3)pLysE which were picked were able to produce hSP56 upon induction by 0.4 mM IPTG. Because the hSP56 is a small protein of 56 kDa, it runs at the front of the 15 % polyacrylamide gel. The expression of hSP56 could be checked by comparing the putative clones with a negative control that was not induced by IPTG (Fig 3-8).

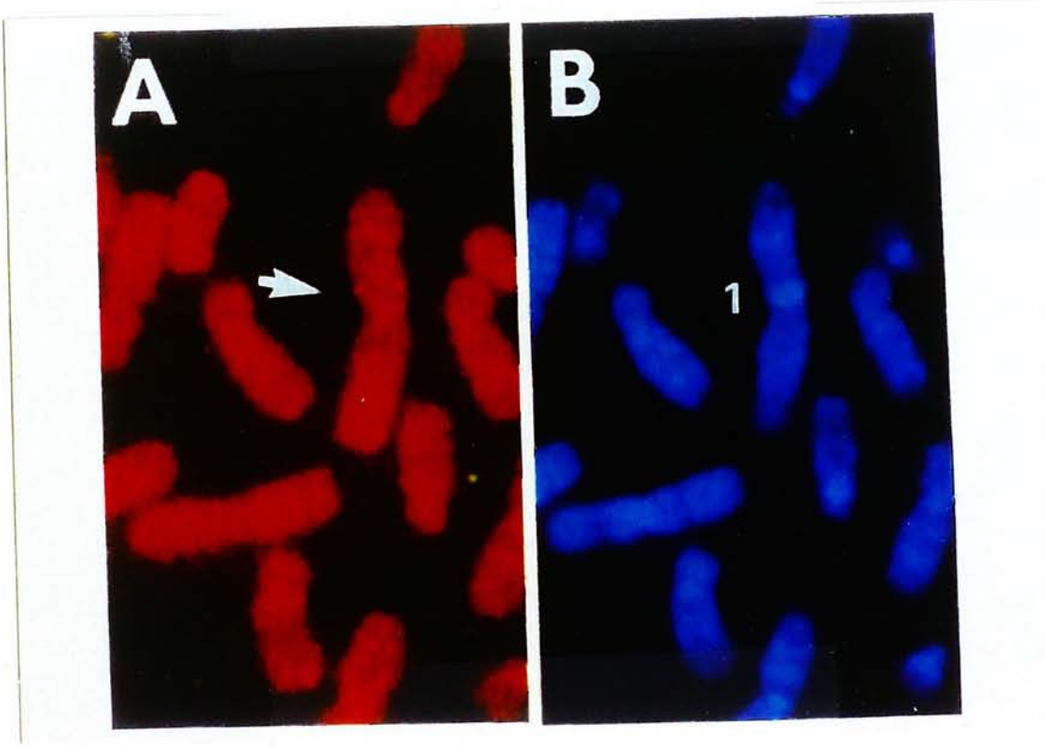
Fig. 3-8. Expression of hSP56 in E. coli BL21(DE3)pLysE. SDS-PAGE of hSP56 with 15 % polyacrylamide gel. Lane 1: uninduced recombinant bacterial crude extract. Lane 2: induce recombinant bacterial crude extract.



3.7 Chromosomal mapping of the hSP56 gene

Under the condition used, the hybridization efficiency was approximately 80% for the hSP56 probe (among 100 checked mitotic figures, 81 of them showed signals on one pair of the chromosomes). Since the DAPI banding was used to identify the specific chromosome, the assignment between the signal from probe and the long arm of chromosome 1 was obtained. The detailed position was further determined based on the summary from 10 photos (Fig3-9-A). There was no additional locus picked up by FISH detection under the condition used, therefore, probe hSP56 is located at human chromosome 1, region q21-q22. No significant FISH signals were observed from other chromosomes. Fig3-9-B shows the mapping results.

Fig. 3-9.(A and B) FISH mapping. A: FISH signals on chromosomes one. B: The same mitotic figures stained with DAPI to identify chromosome 1



CHAPTER 4

DISCUSSION

4.1 General discussion

I have sequence 553 adult human heart cDNA clones. I assigned putative identities of these clones and set up the database and catalogued the results (Table 3-1,3-2,3-3). These are parts of sequencing data in our cDNA library data pool of our group.

The catalogue of those 553 cDNA clones may be useful for the analysis in future. For example, studies of the frequency of cDNA clones that matched to known genes can act as an alternative of Northerning blotting (Davies, 1993). If we know which genes are expressed in the adult human heart, we can estimate the level of expression (Tsui Ph.D. thesis 1995).

Also the categorization of these clones (Table. 3-1) can form an expression profile according to their function. It

could be used as a reference for a human heart tissue when it compared with that other human tissue. When the profile is compared with that of human brain, some interesting results can be obtained (Adams *et al.,* 1993). The percentage of matches that belong to transcription and translation can be used as a reference group because we can assume that the protein synthesis machinery is comparable in heart and brain. (Tsui Ph.D. thesis 1995).

On the other hand, the comparisons of the compositions and expression profiles between normal human heart cDNA library and human diseased heart cDNA libraries can also give valuable data on the change in gene expression in the diseases state. It may help the studying of human heart diseases macroscopically and may lead to breakthroughs in the study of multigenic human heart diseases.

There are 196 clones that belong to human-match genes. Only 11 clones have been match to nonhuman genes. Similarities to nonhuman genes enable us to speculate on

their function in human (Tugendreich *et al.,* 1993). These nonhuman match clones can be used as probes to screen the cDNA library for full length clones by hybridization (Bloem *et al.,* 1990) or PCR (Amaravadi *et al.*, 1994). Also , the full length cDNA can be generated using human heart mRNA as the starting material (Apte *et al.*, 1993; Barnard *et al.*, 1994; Weis, 1994). Other EST sequencing groups have used their nonhuman matches to clone the human homologs of nonhuman proteins (Hori *et al.*, 1994). There are 346 cDNA clones that do not match to any known gene and these may contain inserts encoding novel genes. These genes may be used as the resource for studying gene expression in the human heart and for the discovery of new genes.

Beside the sequencing of 553 cDNA clones, my research project focused on the isolation and partial characterization of a cDNA encoding the hSP56 protein. To date, this is the first selenium binding protein reported in *Homo Sapiens.* Within the cDNA sequence of hSP56, there are no in-frame TGA codons which represent the amino acid selenocysteine [Chamber *et al.,*

1986]. Therefore, hSP56 may interact with selenium in a manner different from the selenoproteins glutathione perioxidase and formate dehydrogenase. This result is in accordance to that of mouse SP56 [Bansal *et al.*, 1990]. Although the selenium binding proteins have not been show to have any specific biochemical functions directly, there is evidence that selenium binding proteins may, under some conditions, perform a functional role for a short time in selenium deficient animals (Walchulewski *et al.*, 1988). Concerning the DNA and amino acid sequence, the putative hSP56 protein is similar to both mouse SP56 and AP56, with a slightly higher similarity to the former. The DNA sequences similarity of hSP56 with those of mSP56 and AP56 are 87.2% and 86.6% respectively. The amino acid sequence of hSP56 have 98.7% similarity and 87.3% identity with mSP56, and 97.9% similarity and 86.4% identity with AP56. It thus seems that hSP56 protein is the human homolog of mouse SP56.

The results for the tissue distribution of the homolog of hSP56 in mouse tissues are very similar to that reported

previously for mSP56 and AP56 whose expressions in liver, kidney and lung are the highest [Lanfear *et al.,* 1993]. However, we further reported that significant levels of SP56 are present in heart. This finding may help to further elucidate the function of this protein.

At present, the physiological role of hSP56 is unknown. Evidence has been presented previously for a putative role of a 58 kDa selenoprotein in growth control in mammalian cells [Morrison *et al.*, 1988]. When the mouse cDNA of SP56 was cloned and sequenced, it was also considered to be a growth regulatory protein [Bansal *et al.,* 1990]. However, the fact that mSP56 and AP56 cannot be detected in mammary cells or mammary cell lines [Lanfear *et al.,* 1993] implies that neither mSP56 nor AP56 is the 58 kDa selenoprotein detected in mammary cells. A 58 kDa selenium binding protein purified from mouse mammary epithelial cells (MMEC) was used to examine wether selenium modulates protein synthesis or is just a marker for cellular selenium status. Previous study results suggest that selenium is attached to the 58 kDa

protein, but does not regulate either its protein synthesis or its functional activity (Sinha *et al.*, 1993). Based on the similarity of protein sequences, we believe that mSP56 and AP56 are functionally similar. Various hypotheses have been proposed on the role of AP56 and SP56 in acetaminophen-induced hepatoxicity: AP56 and SP56 may have important functions which are inhibited by acetaminophen-binding (ultimately leading to cell death); alternatively, AP56 may have a protective role as a scavenger of toxic electrophiles or oxidant species such as acetaminophen metabolites [Pumford et al., 1992; Lanfear *et al.*, 1993]. Previous reports have suggested that acetaminophen detoxification may depend on the selenium status of the animal, since acetaminophen-induced hepatotoxicity and lipid peroxidation seem to be decreased by selenium administration [Schnell *et al.*, 1988; Wendel *et al.*, 1981]. Most recently, Burk *et al.* [1995] reported that selenoprotein P, another selenium binding protein, mediates the protective effect of selenium supplement treatment on liver necrosis induced by free radicals in selenium deficient rat. We report here that hSP56 is present in mouse heart in

significant amounts. It is of interest to test whether hSP56 can mediate the protective effect of selenium supplement treatment against free-radicals-induced necrosis of heart tissue in selenium deficient animals. Clinically, deficiency of selenium can cause heart failure [Yang *et al.*, 1984] and the diseased condition can be alleviated by selenium supplementation. It will be of potential clinical significance to investigate whether there is any protective role of hSP56 throughout the selenium supplement treatment of the selenium deficient patients.

4.2 The possible roles of hSP56 and mSP56

The function of hSP56 is still unknown. mSP56 and hSP56 may have a similar function because they are highly homologous in protein sequences. From the previous study disscussed in section 4.1, we know that one of the selenium-binding protein, selenoprotein P has the protective effect of selenium supplement treatment on liver necrosis induced by free radicals in selenium deficient rat (Bansal *et al.,* 1990; Yang *et al.,* 1984). We are interested to find out whether hSP56 or mSP56 has the same function as the selenoprotein P. At the moment, information is so scarce that it is difficult to make new suggestions regarding the function of the hSP56.

4.3 Further aspects

Antibodies against hSP56 have been produced in our laboratory but whether these peptide-induced antibody is specific for hSP56 or not remains to be determined. In addition to studying the function of hSP56, a good antibody is necessary for the purification of the native protein using either the prokaryotic or the eukaryotic expression system. We can also use a two hybrid system to study whether hSP56 has any interaction with the proteins.

REFERENCES

Adams M.D., Kelley J.M., Gocayne J.D., Dubnick M., Polymeropoulos M.H., Xiao H., Merril C.R., Wu A., Olde B., Moreno R.F., Kerlavage A.R., McCombie W.R. and Venter J.C. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252: 1651-1656.

Adams M.D., Dubnick M., Kerlavage A.R., Moreno R.F., Kelley J.M., Utterback T.R., Nagle J.W., Field C. and Venter J.C. (1992) Sequence identification of 2,375 human brain genes. *Nature* 355: 632-635.

Adams M.D., Kerlavage A.R., Field C. and Venter J.C. (1993) 3,400 new expresseed sequence tags identify diversity of transcripts in human brain. *Nature genet.* 4:256-267.

Adams M.D., Soares M.B., Kerlavage A.R., Fields C. and Venter J.C. (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nature genet.* 4: 373-380.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. *J Mol Biol* 215: 403-410.

Amaravadi L. and King M.W. (1994) A rapid and efficient, nonradioactive method for screening recombinant DNA libraries. *BioTechniques* 16: 98-103.

Ansorge W., Sproat B.S., Stegemann J., Schwager C. and Zenke M. (1987) Automated DNA sequencing: ultrasensitive detection of fluorescent bands during electrophoresis. *Nucleic Acids Res.* 15: 4593-4602.

Apte A.N. and Siebert P.D. (1993) Anchor-ligated cDNA libraries: A technique for generating a cDNA library for the immediate cloning of the 5' ends of mRNAs. *BioTechniques* 15: 890-893.

Bainbridge B.W. (1991) Microbiological techniques for molecular biology: bacteria and phage, *in Essential Molecular Biology: A Practical Approach*, T. A. Brown ed. pp. 13-45, Oxford:IRL Press.

Bairoch A. (1993) The PROSITE dictionary of sites and patterns in proteins, its current status. *Nucleic Acid Res.* 21:3097-3103.

Bansal MP, Cook RG, Danielson KG, Medina D (1989a): A 14-kilodalton selenium-binding protein in mouse liver is fatty acid-binding protein. *J Biol Chem* 264:13780-13784.

Bansal MP, Oborn CJ, Danielson KG, Medina D (1989b): Evidence for two selenium-binding proteins distinct from glutathione peroxidase in mouse liver. *Carcinogenesis* 10: 541-546.

Bansal MP, Mukhopadhyay R, Scott J, Cook RG, Mukhopadhyay R, Medina D (1990): DNA sequencing of a mouse liver protein that binds selenium: implications for selenium's mechanism of action in cancer prevention. *Carcinogenesis* 11: 2071-2073.

Barnard R.B., Southard J.N. and Talamantes F. (1994) Two-step PCR amplification of mutiple specific products from cDNA using one specific primer and oligo dT. *BioTechniques* 16: 251-252.

Bartolone JB, Birge RB, Bulera SJ, Bruno MK, Nishanian EV, Cohen SD, Khairallah EA (1992): Purification, antibody production and partial amino acid sequence of the 58 kDa acetaminophen-binding liver protein. *Toxicol Applied Pharmocol* 113: 19-29.

Bellis G.D., Manoni M., Pergolizzi R., Vezzoni P. and Luzzana M. (1990) Primer design in fluorescent DNA sequencing. *Nucleic Acids Res.* 18: 4951.

Benson D.A., Boguski M., Lipman D.J. and Ostell J. (1993) GenBank. *Nucleic Acid Res.* 22: 3441-3444.

Benson D., Lipman D.J. and Ostell J. (1993) GenBank. *Nucleic Acid Res.* 21: 2963-2965.

Berry T (1994): A selenium transport protein model of a sub-type of schizophrenia. *Med Hypotheses* 43: 409-416.

Bloem L.J. and Yu L. (1990) A time-saving method for screening cDNA or genomic libraries. *Nucleic Acids Res.* 18: 2830.

Bork P., Ouzounis C., Sander C., Scharf M., Schneider R. and Sonnhammer E. (1992) What's in a genome? *Nature* 358: 287.

Burk RF (1983): Biological activity of Selenium. *Ann Rev Nutr* 3: 53-70.

Burk RF.(1995): Selenium in Biology and Human Health. *Springer-Verlag.* pp181-198.

Burk RF, Hill KE, Awad JA, Morrow JD, Kato T, Cockell KA, Lyons PR (1995): Pathogenesis of Diquat-induced liver Necrosis in selenium-deficient rats: assessment of the roles of lipid peroxidation and selenoprotein P. *Hepatology* 21: 561-569.

Chae K.S.. Murakawa K., Okubo K. and Matsubara K. (1994) The use of overlapping and tailed short primers in the chromosomal assignment of short cDNAs by the polymerase chain reaction. *Gene* 142: 199-205.

Chambers I, Frampton J, Goldfarb P, Affara N, McBain W, Harrison PR (1986): The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the termination codon, TGA. *EMBO J* 5: 1221-1227.

Chomczynski P, and Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem.* 162: 156-159.

Cohen M.M. and Emanuel B.S. (1994) Expressed sequence tags. *Science* 266:

Collet C. and Joseph R. (1994) The identification of nuclear and mitochondrial genes by sequencing randomly chosen clones from a marsupial mammary gland cDNA library. *Biochem. genet* 32: 181-190.

Collins F. and Galas D. (1993) A new five-year plan for the U.S. Human Genome Project. *Science* 262: 43-46.

Connell C., Fung S., Heiner C., Bridgham J., Chakerian V., Heron E., Jones B., Menchen S., Mordan W., Raff M., Recknor M., Smith L., Springer J., Woo S. and Hunkapiller M. (1987) Automated DNA sequence analysis. *BioTechniques* 5: 342-348.

Davies K. (1993) The EST express gathers speed. *Nature* 364: 554.

Dorfman D.M. (1993) Amplification of bacteriophage library inserts using polymerase chain reaction. *Method Enzymol.* 218: 336-340.

Dorit R.L., Ohara O. and Gilbert W. (1993) One-sided anchored polymerase chain reaction for amplification and sequencing of complementary DNA. *Method Enzymol.* 218: 36-47.

Durkin A.S., Maglott D.R. and Nierman W.C. (1992) Chromosomal assignment of 38 human brain expressed sequence tags (ESTs) by analyzing fluorescently labeled PCR products from hybrid cell panels. *Genomics* 14: 808-810.

Edwards C.R.W, Bouchier I.A.D, Haslett C, Chilvers E.R. (1996): Davidson's principles and practice of medicine. pp 191-194.

Emard J.F, Thouez J.P, Gauvreau D (1995): Neurodegenerative diseases and risk factors: a literature review. *Soc Sci Med* 40: 847-58.

Frankenberger W.T and Sally Benson J.R (1994): Selenium in the Environment. *Marcel Dekker, Inc.* pp 300-316.

Grothues D., Voss H., Stegemann J., Wiemann S., Sensen C., Zimmermann J., Schwager C., Erfle H., Rupp T. and Ansorge W. (1993) Separation of up to 1000 bases on a modified A.L.F. DNA sequencer. *Nucleic Acids Res.* 21: 6042-6044.

Gubler U. and Hoffman B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263-269.

Heng HHQ, Squire J, Tsui LC (1992): High resolution mapping of mammalian genes by *in situ* hybridization to free chromatin. *Proc Natl Acad Sci USA* 89: 9509-9513.

Heng HHQ, Tsui LC (1993): Modes of DAPI banding and simultaneous *in situ* hybridization. *Chromosoma* 103: 325-332.

Henikoff S. (1993) Sequence analysis by electronic mail server. *Trends Biochem. Sci.* 18:267-268.

Hill KE, Lloyd RS, Yang JG, Read R, Burk RF (1991): The cDNA for rat selenoprotein P contains ten TGA codons in the open reading frame. *J Biol Chem* 266: 10050-10053.

Hill KE, Lloyd RS, Burk RF (1993): Conserved nucleotide sequences in the open reading frame and 3'-untranslated region of selenoprotein P mRNA. *Proc. Natl. Acad. Sci. USA* 90: 537-541.

Hori N., Chae K.S., Murakawa K., Matoba R., Fukushima A., Okubo K. and Matsubara K. (1994) A human cDNA sequence homologue of bovine phosphatidylethanolamine-binding protein. *Gene* 140: 293-294.

Hunkapiller T., Kaiser R.J., Koop B.F. and Hood L. (1991) Large-scale and automated DNA sequence determination. ***Science*** **254: 59-67.**

Hurd DL and Kipling JJ (1964): Origins and growth of Physical Science, part 2, Penguin Books, Harmondsworth.

Hwang D.M., Hwang W.S., Liew C.C. (1994): Single pass sequencing of a unidirectional human fetal heart cDNA library to discover novel genes of the cardiovascular system. ***J Mol Cell Cardiol*** **26: 1329-1333.**

Hwang D.M., Fung YW, Wang RX, Laurenssen CM, Ng SH, Lam WY, Tsui KW, Fung KP, Waye M, Lee CY, Liew CC (1995): Analysis of expressed sequence tags from a fetal human heart cDNA library. ***Genomics*** **30: 293-298.**

Hwang D.M., Fung Y.W., Tsui K.W., Ng S.H., Laurenssen C., Wang R.X., Li Y., Lee C.Y. and Liew C.C. (1994) Gene expression patterns in adult and fetal human hearts analyzed by expressed sequence tags (ESTs) ***Abstract in The Sixth International Genome Sequencing and Analysis Conference*****, Sept. 17-21, South Carolina.**

Hyder S.M., Hu C., Needleman D.S., Sonoda Y., Wang X.Y. and Baker V.V. (1994) Improved accuracy in direct automated DNA sequencing of small PCR products by optimizing the template concentration. *BioTechniques* 17: 478-482.

Khan S.K., Wilcox A.S., Polymeropoulos M.H., Hopkins J.A., Stevens T.J., Robinson M., Orpana A.K. and Sikela J.M. (1992) Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature genet.* 2: 180-185.

Knekt P, Aromaa A, Maatela J, Alfthan G, Aaran RK, Hakama M, Hakulinen T, Peto R, Teppo L (1990): Serum selenium and subsequent risk of cancer among Finnish men and women. *J Natl Cancer Inst* 82: 864-868.

Koshland D.E. (1994) Molecule of the year: The DNA repair enzyme. *Science* 266: 1925.

Lam W.Y., Tsui S.K.W., Lee C.Y. and Waye M.M.Y. (1993) Computer programmes for DNA and protein sequence analysis using an electronic mail server and an IBM personal computer. Abstract for the Second International Conference on Intelligent systems for molecular biology (*ISMB*-94)

Lam W.Y., Tsui S.K.W., Waye M.M.Y., Cheung H.K.Y., Yam N.Y.H., Lo A.S.Y., Law P.T.W. and Lee C.Y. (1993) Isolation of a novel human cDNA coding for a putative heat shock protein from a human heart cDNA library. *Abstract for the American Society of Cell Biology Meeting in* Dec 1993.

Lanfear J, Fleming J, Walker M, Harrison P (1993): Different patterns of regulation of the genes encoding the closely related 56 kDa selenium- and acetaminophen-binding proteins in normal tissues and during carcinogenesis. *Carcinogenesis* 14: 335-340.

Lehr D (1994): A possible beneficial effect of selenium adminstration in antiarrhythmic therapy. *J Am Coll Nutr* 13: 496-8.

Liew C.C., Sole M.J., Yamauchi-takihara K., Kellam B., Anderson D.H., Lin L. and Liew J.C. (1990) Complete sequence and organization of the human cardiac b-myosin heavy chain gene. *Nucleic Acids Res.* 18: 3647-3651.

Liew CC (1993): A human heart cDNA library- the development of an efficient and simple method for automated DNA sequencing. *J Mol Cell Cardiol* 25: 891-894.

Liew CC, Hwang DM, Fung YW, Laurenssen C, Cukerman E, Tsui S, Lee CY (1994): A catalogue of genes in the cardiovascular system identified by expressed sequence tags (ESTs). *Proc Natl Acad Sci USA* 91: 10645-10649.

MacFerrin KD, Chen L, Terranova MP, Schreiber SL, Verdine GL (1993): Overproduction of protein using expression-cassette polymerase chain reaction. *Method Enzymol* 217: 79-102.

Medina D and Oborn CJ (1981): Differential effects of selenium on the growth of mouse mammary cell *in vitro*. *Cancer Lett* 13: 333-344.

Medina D and Oborn CJ (1984): Selenium inhibition of DNA synthesis in mouse mammary epithelial cell line YN-4. *Cancer Res* 44: 4361-4365.

Medina D, Morrison D, Oborn CJ (1985): Selenium retention and inhibition of cell growth in mouse mammary epithelial cell lines *in vitro*. *Biol Trace Element Res* 8: 19-35.

Medina D, Morrison DG (1988): Current ideas on selenium as a chemopreventive agent. *Pathol Immunopathol Res* 7: 187-199.

Morrison DG, Dishart MK, Medina D (1988): Intracellular 58 kDa selenoprotein levels correlated with inhibition of DNA synthesis in mammary epithelial cells. *Carcinogenesis* 9: 1801-1810.

Morrison DG, Bansal MP, Kittrell F, Medina D (1989): Distribution of selenium-binding proteins in different tissues of the mouse and rat. *In Vivo* 3: 16-172.

Matoba R., Okubo K., Hori N., Fukushima A. and Matsubara K. (1994) The addition of 5'-coding information to a 3'-directed cDNA library improves analysis of gene expression. *Gene* 146: 199-207.

Matsubara K. and Okubo K. (1993) cDNA analyses in the human genome project. *Gene* 135: 265-274.

Nebergall WH, Schmidt FC and Holtzclaw HF (1968) General chemistry, 3 rd edn. Raytheon Education, Boston.

Okubo K., Hori N., Matoba R., Niiyama T., Fukushima A., Kojima Y. and Matsubara K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspect of gene expression. Nature genet. 2: 173-179.

Olson M., Hood L., Cantor C. and Botstein D. (1989) A common language for physical mapping of the human genome. *Science* 245: 1434-1435.

Olson M.V. (1993) The human genome project. *Proc. Natl. Acad. Sci.* USA 90: 4338-4344.

Papadopoulos N., Nicolaides N.C., Wei Y.F., Ruben S.M., Carter K.C., Rosen C.A., Haseltine W.A., Fleischmann R.D., Fraser C.M., Adams M.D., Venter J.C., Hamilton S.R., Petersen G.M., Watson P., Lynch H.T., Peltomäki P., Mecklin J.P., Chapelle A., Kinzler K.W. and Vogelatein B. (1994) Mutation of a *mutL* homolog in hereditary colon cancer. *Science* 263: 1625-1629.

Park Y.S. , Kwak J.M., Kwon O-Y., Kim Y.S., Lee D.S., Cho M.J., Lee H.H. and Nam H.G. (1993) Generation of expressed sequence tags of random root cDNA clones of *Brassica napus* by single-run partial sequencing. *Plant physiol.* 103: 359-370.

Polymeropoulos M.H., Xiao H., Glodek A., Gorski M., Adams M.D., Moreno R.F., Fitzgerald M.G. Venter J.C. and Merril C.R. (1992) Chromosomal assignment of 46 brain cDNA. *Genomics* 12: 492-496.

Pumford NR, Martin RM, Hinson JA (1992): A metabolite of acetaminophen covalently binds to the 56 kDa selenium binding protein. *Biochem Biophy Res Commun* 182: 1348-1355.

Putney S.D., Herlihy W.C. and Schimmel P. (1983) A new tropinin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing. *Nature* 302: 718-721.

Read R, Bellew T, Yang JG, Hill KE, Palmer IS, Burk RF (1990): Selenium and amino acid composition of selenoprotein P, the major selenoprotein in rat serum. *J Biol Chem* 265: 17899-17905.

Rechsteiner M.C. (1991) The human genome project: misguided science policy. *Trends Biochem.Sci.* 16: 455-459.

Reilly C (1996): Selenium in Food and Health. Blackie academic & professional. pp 203-217, pp 257-293.

Reilly D (1951): Robert Boyle and his background. *Journal of Chemical Education*, 28: 178-182.

Rice C.M., Fuchs R., Higgins D.G., Stoehr P.J. and Cameron G.N. (1993) The EMBL data library. *Nucleic Acid Res.* 21: 2967-2971.

Robberecht H, Deelstra H (1994): Factors influencing blood selenium concentration values: a literature review. *J Trace Elem Electrolytes Health Dis* 8: 129-143.

Rosenberg A.I., Lade B.N., Chui D.S., Lin S.W., Dunn J.J. and Studier F.W. (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* 56: 125-135.

Saiki R.K., Bugawan T.L., Horn G.T., Mullis K.B. and Erlich H.A. (1986) Analysis of enzymatically amplified b-globin and HLA-DQa DNA with allele-specific oligonucleotide probes. *Nature* 324: 163-166.

Sambrook J., Fritsch E. F and Maniatis T. (1989) Molecular Cloning: A laboratory manual, 2nd ed., Cold Spring Harbor, Cold Spring Harbor University Press.

Salonen JT, Alfthan G, Huttunen JK, Puska P (1984): Association between serum selenium and the risk of cancer. *Am J Epidemiol* 120: 342-349.

Schnell RC, Park KS, Davies MH, Merrick BA, Weir SW (1988): Protective effects of selenium on acetaminophen-induced hepatoxicity of the rat. *Toxicol Applied Pharmacol* 95: 1-11.

Sikela J.M. and Auffray C. (1993) Finding new genes faster than ever. *Nature genet.* 3: 189-191.

Sinha R, Bansal MP, Ganther H, Medina D (1993): Significance of selenium-labeled proteins for selenium's chemopreventive functions. *Carcinogenesis* 14: 1895-1900.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connel CR, Heiner C., Kent SBH. and Hood LE (1986) Fluorescent detection in automated DNA sequence analysis. *Nature* 321: 674-679.

Studier F W and Moffatt B A (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113-130.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990): Use of T7 polymerase to direct expression of cloned genes. *Method Enzymol* 185: 60-89.

Sudo K, Chinen K and Nakamura Y (1994) 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. ***Genomics*** **24: 276-279.**

Oikawa T, Esaki N, Tanaka H and Soda K. (1991): Metalloselenonein, the selenium analogue of metallothionein: synthesis and characterization of its complex with copper ions. ***PNAS*** **88: 3057-3059.**

Tsui SKW (1995) Gene expression in adult human heart as revealed by random sequencing of cDNA library, Ph.D thesis. The Chinese University of Hong Kong.

Tsui SKW, Fung YW, Law PTW., Lam WY, Waye MMY, Lee CY and Liew CC (1994) Large scale cDNA sequencing to analyze gene expression in human heart. ***J. Mol. Cell. Cardiol.*** **26: CLXXVI.**

Tsui SKW, Waye MMY, Lee CY (1995): Efficient automated large-scale sequencing of unpurified PCR product. ***BioTechniques*** **19: 577-578.**

Tsui S.K.W., Waye M.M.Y. and Lee C.Y. (1995) Effects of agar and agarose on Vent™ DNA polymerase and Taq DNA polymerase. ***BioTechnol. Lett.*** **17: 1-6.**

Tsui SKW, Yam NYH, Lee CY, Waye MMY (1994): Isolation and characterization of a cDNA that codes for a LIM-containing protein which is developmentally regulated in heart. *Biochem Biophys Res Commun* 205: 497-505.

Tugendreich S., Boguski M.S., Seldin M.S. and Hieter P. (1993) Linking yeast genetics to the mammalian genomes: identification and mapping of the human homolog of CDC27 via the expressed sequence tag (EST) data base. *Proc. Natl. Acad. Sci. USA* 90: 10031-10035.

Venter J.C. (1993) Identification of new human receptor and transporter genes by high throughput cDNA (EST) sequencing. *J. Pharm. Pharmacol.* 45(Suppl. 1): 355-360.

Vernie LN (1984) Selenium in carcinogenesis. *Biochim Biophy Acta* 738: 203-217.

Voss H., Schjwager C., Wiemann S., Kristensen T., Duthie S., Olsson A., Erfle H., Stegemann J., Zimmermann J. and Ansorge W. (1989) One-step reaction protocol for automated DNA sequencing with T7 DNA polymerase results in uniform labelling. Meth. *Mol. Cell. Biol.* 3: 153-155.

Voss H., Zimmermann J., Schjwager C., Erfle H., Stegemann J.,Stucky K. and Ansorge W. (1990) Automated fluorescent sequencing of lambda DNA. *Nucleic Acids Res.* 18: 5314.

Voss H., Wiemann S., Zimmermann J., Schjwager C., Stegemann J., Erfle H., Rupp T. and Ansorge W. (1993) Automated low-redundancy large scale sequencing by primer walking. *BioTechniques* 15: 714-721.

Virtamo J, Valkeila E, Alfthan G, Punsar s, Huttunen JK, Karvonen MJ (1987): Serum selenium and risk of cancer: A prospective follow-up of nine years. *Cancer* 60: 145-148.

Walchulewski I.H. and Sunde R.A. (1988) Effect of dietary methionine on utilization of tissue selenium from dietary selenomethionine for glutathione peroxidase in the rat. *Journal of Nutrition* 118: 367-374.

Waye M.M.Y., Cheung H.K.Y., Lam W.Y., Law P.T.W., Lo A.S.Y., Lui V.W.Y., Luk S.C.W., Tsui S.K.W., Tung C.K.C., Yam N.Y.H., Liew C.C. and Lee C.Y. (1995) Gene expression of adult human heart as revealed by random sequencing of cDNA library, in Miami Winter BioTechnology Symposium proceedins. Advances in Gene Technology: Protein engineering and structural biology. Cambridge university press. 6: 90.

Weis J.H. (1994) '*Race no more*': an alternative approach to cloning the 5' end of transcripts. *Nucleic Acids Res.* 22: 3427-3428.

Wendel A, Feuerstein S (1981): Drug-induced lipid peroxidation in mice. 1. Modulation by mono-oxygenase activity, glutathione and selenium status. *Biochem Pharmacol* 30: 2513-2520.

Wickens M (1990) How the messenger got its tail: addition of poly (A) in the nucleus. *TIBS* 15: 277-281.

Wilcox A.S., Khan A.S., Hopkins J.A. and Sikela J.M. (1991) Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implication for an expression map of the genome. *Nucleic Acids Res.* 19: 1837-1843.

Willett WC, Stampfer MJ (1986): Selenium and human cancer: Epidemiological aspects and implications for clinical trials. *J Am Coll Toxicol* 5: 29-36.

Yang G, Chen J, Wen Z, Ge K, Zhu L, Chen X, Chen X. (1984): The role of selenium in Keshan disease. *Adv Nutr Res* 6: 203-231.

Yang JG, Morrison-Plummer J, Burk RF (1987): Purification and quantitation of a rat plasma selenoprotein distinct from glutathione peroxidase using monoclonal antibodies. *J Biol Chem* 262: 13372-13375.

Yu W (1988): Collected works of a comprehensive scientific survey on Keshan disease in Chuxiong prefecture. People's Medical Publishing House. (in Chinese) Beijing. pp 21-28.

Zhang D.X. and Hewitt G.M. (1994) *Taq* cycle sequencing is more sensitive to DNA base bias. *BioTechniques* 16: 573-575.

Zimmermann J., Wiemann S., Voss H., Schwager C. and Ansorge W. (1994) Improved fluorescent cycle sequencing protocol allows reading nearly 1000 bases. BioTechniques 17: 302-307.

APPENDIX 1

primer list

1. λgt22-forward PCR primer:

   5'-ATT GGT GGC GAC GAC TCC TGG-3'

2. λgt22-reverse PCR primer:

   5'-TTT GAC ACC AGA CCA GGT A-3'

3. λgt22-forward fluorescent sequencing primer:

   5'-XGGT GGC GAC GAC TCC TGG AGC C-3'

4. λgt22-reverse fluorescent sequencing primer:

   5'-XGAC ACC AGA CCA ACT GGT AAT G-3'

5. hSP56-ATG primer with *Nde*I restriction site:

   5'-TAG GGC CAT ATG GCT ACG AAA TGT GGG AAT TG-3'

6. Oligo-dT primer with *Eco*RI restriction site:

   5'-TAG GGC GAA TTC TTT TTT TTT TTT TTT-3'

7. T7 forward primer with fluorescent label:

   5'-XAAT ACG ACT CAC TAT AG-3'

8. hSP56 mid primer with fluorescent label

8(a) 5'-TGC ACA AGG TCA TTG AGC CCA-3'
nucleotides: (368-388) forward primer.

8(b) 5'-TGC CCC CCA AGA AAG TGA AGG-3'
nucleotides: (892-913) forward primer.

8(c) 5'-CCT TCA CTT TCT TGG GGG GCA-3'
nucleotides: (892-913) reverse primer.

8(d) 5'-CCA GAT GTC AGA GCT ACA ATC-3'
nucleotides: (1408-1428) forward primer.

8(e) 5'-GAT AGC ACC AAG TCG CGC AAC-3'
nucleotides: (270-291) forward primer.

8(f) 5'-GTA TGG GAC TGG CAG CGC CAT-3'
nucleotides: (704-723) forward primer.

8(g) 5'-ATC TGG ATT TGA ACT CCA CCC TCA TCA-3'
nucleotides: (1422-1449) reverse primer.

8(h) 5'-GAT GTA GAC ACA GTA AAA GGA GGG-3'
nucleotides: (1297-1320) forward primer.

APPENDIX 2

The list of 553 cDNA sequences which are novel sequences (novel) or match with partial cDNA sequences (partial). For the clone number, the first alphabet and the following three numbers indicate the identity of the clone in our library and the "F" indicateds the cDNA clones were sequenced by using forward primer. PUT-ID indicateds the putative identity of the clone. ACCESSION and PERCENT-ID indicated the name of the database and the accession number of the nucleotide sequence that matched with the highest score and the matching percentage. LENGTH indicateds the number of base pairs sequenced.

CLONE	PUT_ID	ACCESSION	LENGTH	PERCENT_ID
T001-F	Human Csa-19	gb|U12404|	266	100
T002-F	Novel			
T003-F	Human titin	emb|X90568|	218	97
T004-F	Novel			
T006-F	Human calpastatin	gb|U31345|	253	98
T009-F	Human acidic ribosomal phosphoprotein P0	gb|M17885|	250	97
T011-F	Novel			
T013-F	Alu repeats			
T014-F	Novel			
T016-F	Human mitochondrial genome			
T019-F	Novel			
T020-F	Human mitochondrial genome			
T022-F	Human mitochondrial genome			
T025-F	Human ATP-citrate lyase	emb|X64330|	174	100
T026-F	Human mitochondrial genome			
T027-F	Human myosin heavy chain, cardic beta	gb|M25137|	170	78
T028-F	Novel			
T030-F	Human NADH-ubiquinone oxidoreductase	gb|M22538|	236	99
T031-F	Human mitochondrial genome			
T033-F	Novel			
T035-F	Novel			
T036-F	Rat Enigma	gb|U48247|	210	90
T038-F	Novel			
T039-F	Novel			
T040-F	Novel			
T042-F	Human myosin light chain 2, ventricular	gb|S69022|	160	95
T045-F	Human mitochondrial genome			
T048-F	Human mitochondrial genome			
T051-F	Human elongation factor 1-alpha	gb|L10340|	194	97
T052-F	Human serine/threonine kinase receptor	gb|U20165|	193	99

T054-F	Novel			
T057-F	Novel			
T058-F	Human glycogen 4-alpha-D-glycosytransferase	gb|U32573|	415	91
T059-F	Human mitochondrial genome			
T061-F	Human myosin light chain 2, cardiac	gb|M22815|	380	99
T062-F	Human mitochondrial genome			
T063-F	Human mitochondrial genome			
T064-F	Human cardiodilation atrial natriuretic factor	gb|I01397|	192	99
T065-F	Human heat shock protein, 71 kDa	emb|Y00371|	304	96
T067-F	Novel			
T069-F	Repetitive sequence			
T071-F	Human mitochondrial genome			
T074-F	Human mitochondrial genome			
T076-F	Human myosin light chain 2, ventricular	gb|S69022|	260	92
T078-F	Human ribosomal protein L30	gb|M94314|	236	98
T080-F	Human myosin light chain 1, ventricular	emb|X07373|	247	97
T080-F	Human mitochondrial genome			
T081-F	Human mitochondrial genome			
T083-F	Novel			
T086-F	Human mitochondrial genome			
T091-F	Human prostaglandin D synthase	gb|M61900|	215	92
T092-F	Novel			
T094-F	Human cardiodilation atrial natriuretic factor	gb|M30262|	334	97
T095-F	Human desmin	gb|M63391|	390	100
T102-F	Human ferritin heavy chain	gb|M12937|	282	98
T107-F	Novel			
T115-F	Human serine/threonine kinase receptor	gb|U20165|	294	100
T117-F	Novel			
T132-F	Human Tob	dbj|D38305|	265	100
T133-F	Human troponin T, cardiac	gb|L40162|	263	100
T141-F	Human protein kinase C inhibitor-I	gb|U27143|	452	98
T143-F	Human plasma protein S, vitamin K-dependent	gb|M15036|	413	99
T147-F	Human metallothionein I-A	gb|K01383|	250	89
T151-F	Human ras-like protein	gb|M31468|	278	92
T153-F	Human heat shock protein, 90 kDa	gb|M16660|	300	86
T160-F	Human BTG1	emb|X61123|	316	99
T162-F	Human myosin light chain 2, cardiac	gb|M22815|	270	98
T164-F	Human mitochondrial genome			
T165-F	Ribosomal RNA			
T166-F	Novel			
T168-F	Human mitochondrial genome			
T170-F	Repetitive sequence			
T171-F	Novel			
T172-F	Human mitochondrial genome			
T176-F	Novel			
T177-F	Novel			
T178-F	Novel			
T183-F	Novel			
T184-F	Novel			
T185-F	Human mitochondrial genome			
T187-F	Human F1-ATPase beta subunit	emb|X03559|	206	81
T189-F	Human mitochondrial genome			
T192-F	Human desmoplakin	gb|J05211|	165	96
T193-F	Novel			
T195-F	Human mitochondrial genome			
T197-F	Human mitochondrial genome			
T198-F	Human mitochondrial genome			
T199-F	Human mitochondrial genome			
T200-F	Human ribosomal protein L4	dbj|D23660|	266	96
T201-F	Human cathepsin L	gb|L06426|	190	82
T202-F	Rat calsequestrin	gb|U33287|	390	83
T203-F	Human myosin heavy chain, cardic beta	gb|M21665|	220	93
T204-F	Human pyruvate kinase, M-2 type	gb|M23725|	234	93
T205-F	Human desmin	gb|M63391|	280	99
T206-F	Human mitochondrial genome			
T208-F	Ribosomal RNA			
T211-F	Human alpha-2-macroglobulin receptor-associated proteg	gb|M63959	392	96
T213-F	Novel			
T215-F	Novel			
T216-F	Human Ki nuclear autoantigen	gb|U11292|	365	98
T217-F	Human mitochondrial genome			
T218-F	Human creatine kinase, muscle	gb|M21494|	173	99
T223-F	Novel			
T224-F	Novel			
T225-F	Novel			
T226-F	Human manganese superoxide dismutase	emb|X65965|	182	86
T235-F	Novel			
T236-F	Novel			
T247-F	Human mitochondrial genome			
T249-F	Repetitive sequence			
T250-F	Human mitochondrial genome			
T252-F	Novel			
T255-F	Human U1 small nuclear RNP-specific C protein	emb|X12517|	252	96
T256-F	Human creatine kinase, mitochondrial	gb|J05401|	235	100
T260-F	Human mitochondrial genome			
T262-F	Rat integral membrane glycoprotein	emb|Z21513|	310	87
T263-F	Novel			
T264-F	Human heat shock protein, 70B'	emb|X51758|	266	99
T265-F	Novel			
T267-F	Novel			

T271-F	Novel			
T272-F	Human ETS2	emb|X55181|	208	95
T273-F	Novel			
T285-F	Human myosin regulatory light chain	emb|X54304|	184	96
T299-F	Human calcium-ATPase	gb|M23115|	151	94
T303-F	Novel			
T307-F	Novel			
T313-F	Human metallothionein I-G	gb|J03910|	295	97
T315-F	Human mitochondrial genome			
T318-F	Human mitochondrial genome			
T319-F	Novel			
T321-F	Human ribosomal protein L26	emb|X69392|	294	96
T327-F	Human metallothionein II	gb|M26637|	170	95
T337-F	Human mitochondrial genome			
T338-F	Novel			
T345-F	Human cytochrome c oxidase, subunit VIIc	emb|X16560|	160	97
T351-F	Human mitochondrial genome			
T352-F	Human glutamate receptor flip isoform	gb|U10301|	163	94
T353-F	Human mitochondrial genome			
T372-F	Novel			
T375-F	Novel			
T376-F	Novel			
T380-F	Human ribosomal protein L26	emb|X69392|	160	95
T386-F	Human mitochondrial genome			
T390-F	Human catenin, alpha	dbj|D13866|	179	97
T393-F	Novel			
T416-F	Novel			
T417-F	Human actinin, skeletal alpha-2	gb|M86406|	212	87
T423-F	Novel			
T424-F	Human brain natriuretic protein	gb|M25296|	279	99
T425-F	Novel			
T439-F	Repetitive sequence			
T442-F	Human troponin T, cardiac	emb|X79857|	264	96
T449-F	Human titin	emb|X90568|	347	99
T450-F	Human titin	emb|X90568|	315	97
T461-F	Human ribosomal protein S5	gb|U14970|	239	96
T466-F	Human tropomyosin, skeletal alpha	gb|M19713|	395	96
T469-F	Human titin	emb|X90568|	293	98
T472-F	Human titin	emb|X90568|	269	96
T475-F	Human mitochondrial genome			
T478-F	Human mitochondrial genome			
T481-F	Human mitochondrial genome			
T484-F	Human mitochondrial genome			
T485-F	Novel			
T487-F	Human mitochondrial genome			
T488-F	Repetitive sequence			
T489-F	Ribosomal RNA			
T493-F	Novel			
T495-F	Novel			
T496-F	Novel			
T503-F	Human tropomyosin, skeletal alpha	gb|M19714|	269	98
T504-F	Human mitochondrial genome			
T505-F	Human myosin heavy chain, cardic beta	emb|X05631|	240	98
T507-F	Novel			
T508-F	Novel			
T509-F	Human growth factor inducible intermediate	emb|X56790|	321	75
T532-F	Human mitochondrial genome			
T533-F	Human mitochondrial genome			
T538-F	Human troponin C, slow skeletal	emb|X07897|	194	68
T544-F	Human mitochondrial genome			
T548-F	Human ribosomal protein L37a	gb|L22154|	235	99
T552-F	Human ribosomal protein L27a	gb|U14968|	163	91
T553-F	Human ubiquitin-like protein	dbj|D23662|	215	98
T554-F	Human ribosomal protein L9	gb|U09954|	242	81
T560-F	Human mitochondrial genome			
T561-F	Novel			
T565-F	Human mitochondrial genome			
T566-F	Human cytochrome c oxidase, subunit Va	gb|M22760|	175	85
T567-F	Novel			
T569-F	Human axonal transporter	emb|X90840|	200	76
T572-F	Human mitochondrial genome			
T573-F	Human moesin	gb|M69066|	180	98
T576-F	Human supressor for yeast mutant	dbj|D66904|	201	89
T578-F	Human mitochondrial genome			
T579-F	Human mitochondrial genome			
T580-F	Novel			
T583-F	Novel			
T584-F	Novel			
T585-F	Novel			
T587-F	Human mitochondrial genome			
T588-F	Human mitochondrial genome			
T591-F	Novel			
T594-F	Human myosin regulatory light chain	emb|X54304|	301	86
T595-F	Human mitochondrial genome			
T596-F	Human tumor necrosis factor alpha inducible protein A	gb|M59465|	355	98
T597-F	Human mitochondrial genome			
T598-F	Human mitochondrial genome			
T599-F	Human DGCR6	emb|X96484|	293	95
T600-F	Novel			
T601-F	Human myosin alkali light chain, non-muscle	gb|M22918|	336	98

T602-F	Novel			
T606-F	Human mitochondrial genome			
T607-F	Human mitochondrial genome			
T610-F	Human mitochondrial genome			
T612-F	Ribosomal RNA			
T613-F	Human mitochondrial genome			
T614-F	Human mitochondrial genome			
T616-F	Human calcyclin	gb|J02763|	379	99
T617-F	Novel			
T620-F	Human ribosomal protein S10	gb|U14972|	570	98
T621-F	Human actin, alpha	emb|X00351|	197	100
T623-F	Human hnRNP E1	gb|X78137|	249	97
T626-F	Human metallothionein II	gb|M26637|	305	92
T627-F	Human chondroitin/dermatan sulfate proteoglycan	gb|M14219|	202	99
T628-F	Human membrane glycoprotein 4F2 antigen heavy chain	gb|J02939|	345	97
T629-F	Novel			
T630-F	Novel			
T637-F	Alu repeats			
T638-F	Novel			
T639-F	Ribosomal RNA			
T640-F	Ribosomal RNA			
T641-F	Novel			
T642-F	Novel			
T644-F	Novel			
T645-F	Human desmin	gb|M63391|	153	93
T646-F	Novel			
T648-F	Human zinc finger protein, bcl-6	gb|U00115|	205	98
T652-F	Novel			
T653-F	Alu repeats			
T655-F	Novel			
T656-F	Human troponin T, cardiac	emb|X74819|	350	98
T657-F	Alu repeats			
T658-F	Human long-chain acyl-Co-A synthetase	dbj|D10040|	331	95
T660-F	Human vinculin	gb|M33308|	323	97
T664-F	Human mitochondrial genome			
T665-F	Human pyruvate dehydrogenase E1-alpha subunit	gb|L48690|	325	98
T667-F	Human ribosomal protein S21	gb|L04483|	222	94
T668-F	Novel			
T671-F	Human mitochondrial genome			
T672-F	Novel			
T673-F	Human merosin	gb|M59832|	337	98
T674-F	Human mitochondrial genome			
T675-F	Human ribosomal protein L18	gb|L11566|	289	98
T676-F	Novel			
T677-F	Human myosin heavy chain, cardic beta	gb|M21665|	188	96
T679-F	Human decorin	gb|M98262|	171	95
T680-F	Human ferritin light chain	gb|M11147|	188	98
T681-F	Human homolog of Drosophila enhancer of split	gb|U04241|	174	98
T682-F	Human mitochondrial genome			
T684-F	Cow F1F0-ATP synthase complex F0 membrane domain g	sugb|S70448|	379	89
T686-F	Human cathepsin B	gb|L22569|	360	97
T688-F	Human mitochondrial genome			
T690-F	Novel			
T692-F	Human ferritin heavy chain	gb|M15383|	156	98
T696-F	Human mitochondrial genome			
T698-F	Human mitochondrial genome			
T700-F	Novel			
T701-F	Human troponin T, cardiac	gb|L40162|	384	99
T702-F	Novel			
T704-F	Human mitochondrial genome			
T705-F	Human mitochondrial genome			
T706-F	Human glycosylated surface protein	emb|X06296|	338	99
T707-F	Human mitochondrial genome			
T708-F	Human very-long-chain acyl-CoA dehydrogenase	dbj|D43682|	303	100
T710-F	Human porin	gb|L08666|	316	97
T711-F	Human actin depolymerization factor	gb|S65738|	385	98
T712-F	Novel			
T713-F	Novel			
T714-F	Human mitochondrial genome			
T715-F	Human cardiodilation atrial natriuretic factor	gb|M30262|	295	93
T716-F	Human mitochondrial genome			
T717-F	Novel			
T719-F	Novel			
T720-F	Mouse 60S ribosomal protein	gb|U28917|	218	85
T721-F	Human troponin I, cardiac	emb|X54163|	236	97
T722-F	Human myosin regulatory light chain	emb|X54304|	196	94
T724-F	Novel			
T725-F	Novel			
T726-F	Human troponin C, slow skeletal	emb|X07897|	178	98
T727-F	Human mitochondrial genome			
T728-F	Novel			
T729-F	Human mitochondrial genome			
T731-F	Novel			
T732-F	Human mitochondrial genome			
T734-F	Human mitochondrial genome			
T735-F	Human 26S protease regulatory subunit	gb|L02426|	358	98
T736-F	Human ubiquitin	gb|M26880|	352	98
T741-F	Human DS-1	emb|X81788|	317	98
T743-F	Human mitochondrial genome			
T745-F	Novel			

T746-F	Novel			
T748-F	Human ribosomal protein S4X	gb|M58458|	210	97
T749-F	Novel			
T750-F	Human mitochondrial genome			
T751-F	Novel			
T753-F	Human acidic ribosomal phosphoprotein P2	gb|M17887|	215	98
T754-F	Human mitochondrial genome			
T758-F	Alu repeats			
T760-F	Human cytochrome c oxidase, subunit VIIc	emb|X16560|	257	98
T765-F	Human mitochondrial genome			
T766-F	Human histidine-rich calcium binding protein	gb|M60052|	401	97
T767-F	Novel			
T770-F	Alu repeats			
T771-F	Human nuclear factor p97	gb|L39793|	341	88
T772-F	Human mitochondrial genome			
T773-F	Human myosin heavy chain, cardic beta	emb|X05631|	177	94
T774-F	Human mitochondrial genome			
T775-F	Novel			
T776-F	Novel			
T777-F	Rat vascular endothelial growth factor	gb|U22372|	303	98
T781-F	Novel			
T782-F	Novel			
T784-F	Human ribosomal protein S16	gb|M60854|	330	99
T785-F	Novel			
T786-F	Human mitochondrial genome			
T789-F	Human mitochondrial genome			
T790-F	Human mitochondrial genome			
T791-F	Novel			
T793-F	Novel			
T795-F	Human prolactin receptor-associated protein	gb|M18981|	313	93
T798-F	Human mitochondrial genome			
T801-F	Human mitochondrial genome			
T803-F	Human troponin C, slow skeletal	emb|X07897|	215	98
T804-F	Human mitochondrial genome			
T805-F	Human mitochondrial genome			
T806-F	Novel			
T808-F	Human myosin heavy chain, cardic beta	gb|M21665|	305	98
T811-F	Human interferon-inducible gene 1-8U	emb|X57352|	250	99
T812-F	Human leukemia virus receptor 1	gb|L20859|	256	97
T813-F	Human mitochondrial genome			
T814-F	Human branched chain acyltransferase	gb|J03208|	213	98
T817-F	Human mitochondrial genome			
T818-F	Human metallothionein from cadium-treated cells	emb|V00594|	220	94
T822-F	Novel			
T823-F	Novel			
T825-F	Novel			
T826-F	Human mitochondrial genome			
T828-F	Human mitochondrial genome			
T832-F	Human mitochondrial genome			
T833-F	Human mitochondrial genome			
T834-F	Human myosin light chain 2, ventricular	gb|S69022|	290	95
T835-F	Human mitochondrial genome			
T836-F	Human ubiquitin	gb|M26880|	238	97
T837-F	Human mitochondrial genome			
T838-F	Ribosomal RNA			
T839-F	Human mitochondrial genome			
T840-F	Alu repeats			
T841-F	Human neurodap1	dbj|D32249|	357	91
T842-F	Human mitochondrial genome			
T846-F	Alu repeats			
T848-F	Alu repeats			
T850-F	Novel			
T851-F	Human mitochondrial genome			
T852-F	Human mitochondrial genome			
T853-F	Human mitochondrial genome			
T854-F	Rat protein tyrosin kinase JAK2	gb|U13396|	152	89
T855-F	Human mitochondrial genome			
T856-F	Human mitochondrial genome			
T857-F	Alu repeats			
T861-F	Human gap junction protein, cardiac	emb|X52947|	157	100
T862-F	Novel			
T863-F	Human lipoprotein lipase	emb|X54516|	252	93
T864-F	Human mitochondrial genome			
T865-F	Human ATPase coupling factor 6 subunit, mitochondrial	gb|M37104|	312	95
T870-F	Human homolog of rat HREV107-like protein	emb|X92814|	308	98
T871-F	Novel			
T873-F	Novel			
T875-F	Human thrombospondin-4	emb|Z19585|	210	85
T876-F	Human mitochondrial genome			
T877-F	Human ubiquitin	gb|M26880|	223	
T878-F	Human platelet-endothelial tetraspan antigen 3	gb|U14650|	160	98
T879-F	Novel			
T880-F	Human myosin regulatory light chain	emb|X54304|	417	98
T882-F	Human mitochondrial genome			
T884-F	Novel			
T885-F	Novel			
T886-F	Novel			
T888-F	Human visinin-like peptide 1 homolog	gb|U14747|	260	84
T889-F	Novel			
T892-F	Human mitochondrial genome			

T894-F	Novel			
T896-F	Human mitochondrial genome			
T897-F	Human mitochondrial genome			
T898-F	Novel			
T899-F	Novel			
T900-F	Human mitochondrial genome			
T901-F	Rat nuclear pore complex protein NUP107	gb|L31840|	380	85
T902-F	Human mitochondrial genome			
T903-F	Novel			
T904-F	Human CAC and GTG repeat-containing mRNA	gb|U00943|	246	97
T905-F	Human myosin heavy chain, cardic beta	gb|M21665|	243	97
T906-F	Human mitochondrial genome			
T908-F	Human myosin light chain 2, ventricular	gb|S69022|	236	100
T909-F	Human N-ras	emb|X02751|	240	93
T910-F	Human myoblast cell surface antigen 24	emb|X16850|	222	91
T914-F	Human mitochondrial genome			
T916-F	Human ribosomal protein S14	gb|M13934|	290	99
T917-F	Novel			
T919-F	Novel			
T920-F	Novel			
T923-F	Human metallothionein I-e	gb|M10942|	190	
T925-F	Novel			
T926-F	Human mitochondrial genome			
T928-F	Human nucleoporin NUP358	gb|L41840|	383	100
T929-F	Human cardiodilation atrial natriuretic factor	gb|M30262|	206	98
T930-F	Human mitochondrial genome			
T931-F	Novel			
T936-F	Human mitochondrial genome			
T937-F	Human ribosomal protein S4X	gb|M58458|	188	98
T938-F	Human mitochondrial genome			
T940-F	Human mitochondrial genome			
T942-F	Novel			
T943-F	Human mitochondrial genome			
T944-F	Human ribosomal protein S4X	gb|M58458|	263	100
T948-F	Human mitochondrial genome			
T949-F	Novel			
T950-F	Human mitochondrial genome			
T951-F	Novel			
T953-F	Human mitochondrial genome			
T954-F	Human nicotinamide nucleotide transhydrogenase	gb|U40490|	154	95
T958-F	Human mitochondrial genome			
T961-F	Human mitochondrial genome			
T964-F	Human creatine kinase, mitochondrial	gb|J05401|	157	97
T966-F	Human mitochondrial genome			
T969-F	Novel			
T970-F	Novel			
T972-F	Human mitochondrial genome			
T974-F	Repetitive sequence			
T980-F	Human cytochrome c oxidase, subunit VIIb	emb|Z14244|	176	97
T982-F	Human desmin	gb|M63391|	186	97
T989-F	Human nucleoporin NUP358	dbj|D42063|	186	98
T991-F	Human mitochondrial genome			
T992-F	Repetitive sequence			
T993-F	Novel			
T994-F	Human metallothionein from cadium-treated cells	emb|V00594|	197	100
T995-F	Human titin	emb|X69490|	252	98
T996-F	Repetitive sequence			
T997-F	Human macroglobulin, beta-2	emb|X07621|	168	91
T999-F	Human mitochondrial genome			
U032-F	Human hnRNP C2	gb|M29063|	284	96
U034-F	Human mitochondrial genome			
U036-F	Alu repeats			
U040-F	Human mitochondrial genome			
U042-F	Human tropomyosin, skeletal alpha	gb|M19714|	165	93
U043-F	Human cytochrome bc-1 complex core protein II	gb|J04973|	231	99
U046-F	Human mitochondrial genome			
U047-F	Novel			
U048-F	Cow ATP synthetase, mitochondrial epsilon subunit	emb|M16978|	211	88
U049-F	Human mitochondrial genome			
U050-F	Novel			
U051-F	Human OXA1Hs	emb|X80695|	208	94
U056-F	Human mitochondrial genome			
U057-F	Cow NADH ubiquinone oxidoreductase, MLRQ subunit	emb|X64897|	319	83
U059-F	Human mitochondrial genome			
U060-F	Human mitochondrial genome			
U061-F	Human elongation factor 1-alpha	gb|M29548|	233	95
U062-F	Human mitochondrial genome			
U063-F	Novel			
U064-F	Human mitochondrial genome			
U066-F	Human mitochondrial genome			
U067-F	Human major nuclear matrix protein	gb|M63483|	230	97
U068-F	Novel		283	
U071-F	Mouse inosine-5'-monophosphate dehydrogenase	gb|M98333|	213	98
U072-F	Human actin, cardiac alpha	emb|X03767|	153	95
U074-F	Human cytochrome c oxidase, subunit VIII	gb|J04823|	166	97
U076-F	Novel			
U077-F	Novel			
U078-F	Novel			
U083-F	Human mitochondrial genome			
U087-F	Novel			

U093-F	Human mitochondrial genome			
U097-F	Novel			
U099-F	Novel			
U100-F	Human ribosomal protein L27a	gb|U14968|	238	100
U101-F	Human mitochondrial genome			
U102-F	Human mitochondrial genome			
U104-F	Human mitochondrial genome			
U109-F	Human mitochondrial genome			
U110-F	Human ribosomal protein L30	emb|X79238|	247	93
U112-F	Human skeletal muscle 190 kDa protein	emb|X69090|	171	69
U115-F	Human short chain acyl-Co-A dehydrogenase	gb|M26393|	154	89
U116-F	Novel			
U117-F	Human actin, cardiac alpha	gb|J00068|	242	94
U118-F	Alu repeats			
U120-F	Human mitochondrial genome			
U121-F	Human mitochondrial genome			
U122-F	Human malate dehydrogenase, cytosolic	dbj|D55654|	167	95
U128-F	Human p190-B	gb|U17032|	247	92
U130-F	Human mitochondrial genome			
U131-F	Human mitochondrial genome			
U133-F	Novel			
U134-F	Human cytochrome c oxidase, subunit VIIc	emb|X16560|	166	70
U135-F	Human proteasome subunit HsN3	dbj|D26600|	222	98
U136-F	Novel			
U140-F	Human mitochondrial genome			
U141-F	Human mitochondrial genome			
U146-F	Novel			
U147-F	Novel			
U148-F	Human inositol phospholipid assembly protein	gb|S74936|	310	98
U149-F	Novel			
U150-F	Alu repeats			
U151-F	Human mitochondrial genome			
U153-F	Human ribosomal protein L23a	gb|U02032|	327	96
U154-F	Human mitochondrial genome			
U155-F	Human mitochondrial genome			
U157-F	Human mitochondrial genome			
U159-F	Human ubiquinol cytochrome c reductase core I protein	gb|L16842|	277	98
U160-F	Alu repeats			
U161-F	Novel			
U162-F	Novel			
U164-F	Human carbonic anhydrase III	gb|M29453|	204	96
U165-F	Novel			
U167-F	Alu repeats			
U171-F	Human crystallin, alpha B	gb|S45630|	347	96
U174-F	Human cytochrome c oxidase, subunit VIc	emb|X13238|	197	98
U175-F	Human mitochondrial genome			
U176-F	Human glucose phosphate isomerase	gb|M55538|	167	98
U177-F	Alu repeats			
U181-F	Novel			
U183-F	Human myoglobin	emb|X00373|2	219	99
U185-F	Repetitive sequence			
U194-F	Human ribosomal protein L39	emb|X82551|	194	88
U195-F	Human heat shock protein, 70 kDa	emb|X04676|	270	98
U197-F	Novel			
U198-F	Human myosin alkali light chain, ventricular	gb|M24248|	252	97
U202-F	Human beta-sarcoglycan dystrophin-associated glycoprog	gb|U29586|	192	97
U203-F	Human cytochrome bc-1 complex core protein II	gb|J04973|	239	98
U205-F	Human ribosomal prtoein L9	gb|U21138|	290	97
U206-F	Novel			
U207-F	Human mitochondrial genome			
U208-F	Novel			
U210-F	Human mitochondrial genome			
U211-F	Human ribosomal protein S18	emb|X69150|	177	98
U212-F	Human nucleoporin-like protein	emb|X89478|	220	98
U213-F	Novel			
U214-F	Human tissue inhibitor of metalloproteinases-3	gb|U14394|	164	93
U215-F	Human ferritin heavy chain	gb|M12937|	300	99
U216-F	Human mitochondrial genome			
U216-F	Human highly basic protein, 23 kDa	emb|X56932|	231	96
U219-F	Human mitochondrial genome			
U220-F	Human mitochondrial genome			
U221-F	Human lysyl synthetase	dbj|D31890|	205	99
U222-F	Human mitochondrial genome			